

Time to accept realities of bioweapon control

The costs of preventing the barbarity of biological warfare are not only financial but also include intrusiveness, at home as well as abroad. There is an urgent need to shoulder those costs to ensure an effective ban.

There will inevitably be a temptation to relax a little after what appears to have been the successful resolution of the latest confrontation with President Saddam Hussein of Iraq. If the agreement brokered by the UN Secretary-General, Kofi Annan, holds, the West can increase its confidence in the eventual destruction of Saddam's arsenal of chemical and biological weapons, while Saddam can offer his people the hope that such a move will lead to the lifting of economic sanctions. But relaxation would be a mistake. The Iraq crisis has been a stark reminder that, even though the use of such weapons is outlawed by two international treaties, the threat they represent remains very real.

It is therefore pleasing that President Bill Clinton has recently released a new statement of US policy that addresses one aspect of this threat, namely support for the speedy completion of a protocol to the Biological Weapons Convention of 1972, outlining how compliance with this convention might be monitored and verified (see page 831). Pressure for such a protocol has not come from the recent behaviour of Iraq alone; other countries have been arguing the case for several years, and have set in motion diplomatic negotiations which the United States has now agreed to join. But Saddam Hussein's actions have certainly focused minds in a way that few other events could have done.

The situation in Iraq has helped to clear the air in other ways too. Some of the traditional US scepticism about the value of verification procedures should have been tempered by the ingenuity and effectiveness displayed by United Nations weapons inspectors since the ending of the Gulf War in 1991. Those who had argued that such inspectors would be easily hoodwinked have had to eat their words. And the argument that a verification regime would make little impact has, in

the process, lost much of its weight. The task facing those negotiating a draft protocol, with Clinton's support now ringing in their ears, is to design a regime that is rigorous, effective and trustworthy.

Formidable obstacles remain. The pharmaceutical and biotechnology industries, for example, both in the United States and elsewhere, have raised concerns — some legitimate, some possibly less so — about the potential threat to commercial confidentiality inevitably created by a visit by foreign inspectors. Other major powers, in particular Russia and China, as well as Japan, have to be persuaded to share the political enthusiasm of the West for strict controls. And countries from the poorer 'South' will expect some form of compensation for their support.

But none of these problems is insurmountable. Experience with the Chemical Weapons Convention of 1993 has shown that it is possible to design inspection procedures that the industry does not find excessively burdensome or intrusive. All major governments must be sufficiently worried about the threat of terrorists releasing biological weapons onto their soil to wish to reduce the likelihood of that happening. And possible compensation schemes for the South — such as an international network to monitor emerging diseases — would not necessarily involve a massive transfer of either proprietary technology or intellectual property, two of industry's greatest fears.

The most difficult task will be to persuade Western nations to accept the costs involved, whether the relative loss of sovereignty implied by an international monitoring regime, the extra paperwork and regulations that would entail, or just the expense of an institutional apparatus designed to make it work effectively. But given terrifying scenarios that are now all too possible, those costs will be a small price to pay for the extra security they will buy. □

Cloning's confused signals

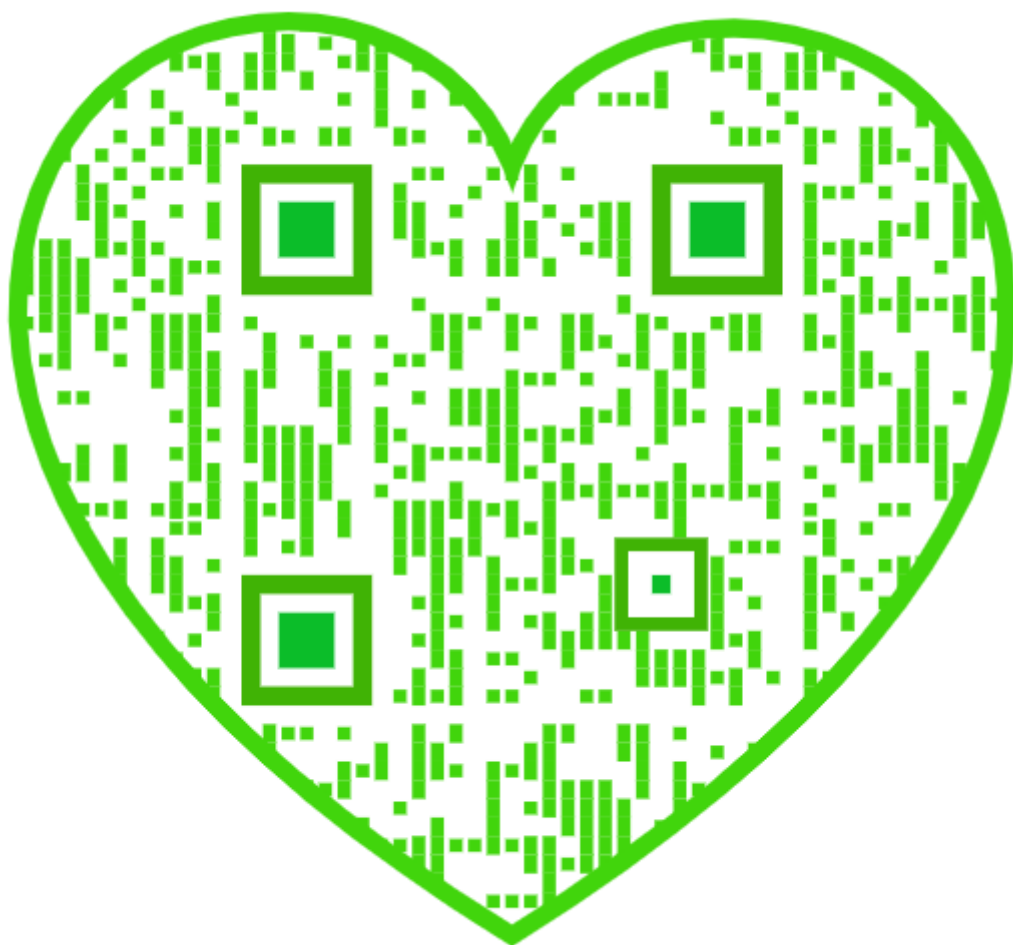
Worries about failure to replicate Dolly are premature but stimulating.

There it was last week, in newspapers and, as follows, high up the CNN Web site: "The Scottish scientist credited with successfully cloning Dolly the sheep last year has admitted he may have made a procedural mistake... 'There is a remote possibility that the cell came from a fetus rather than an adult,' Ian Wilmut said..." Recently in *Science*, an attack on the original *Nature* paper met with a rebuttal from the original authors. But has the public been misled?

No more than in any significant but scientifically controversial development. One controversy concerns the burden of proof. How burdensome must it be, in an area where experiments are both time-consuming and complex? With hindsight (and few could have anticipated the extent of the fuss stimulated by Dolly), it is always easy to claim that the proof was inadequate. But the definition of the threshold of proof is itself controversial, particularly given the significance that the results have taken on. According to soundings across the relevant research community (see page 825), the published attack

represents an extreme position in its demands for proof. And according to encouragingly many, the probability that the paper's results are valid remains high. But this case is yet another illustration of the truism that a formal scientific publication need not necessarily be the last word.

In the meantime, the journals give formalized expression to the buzz of gossip and debate within the community. That debate, as it seeps out, can undermine in the public eye even the most rigorously justified claim. Yet whatever the risks for public perception, it is important that the plurality of scientific views is expressed in the open — which is where informal opinion in news and correspondence pages or on the Web plays its valuable part. It is also important that the public should recognize the current debate for what it is: scientists behaving critically, as they should. Most importantly, it is premature to draw any conclusions from the absence so far of a second mammal cloned from adult cells. □



Dolly researcher plans further experiments after challenges

[PARIS] Ian Wilmut, the Scottish researcher whose group last year reported the cloning of Dolly the lamb from the udder cell of an adult ewe, is planning further experiments to clarify the lamb's origin — and has offered to make frozen tissue from the donor ewe available to other researchers.

The move comes as many embryologists around the world have rallied to Wilmut's defence in rejecting challenges to Dolly's authenticity and playing down the significance of criticisms that his paper (*Nature* 385, 810; 1997) lacked unequivocal evidence that Dolly was derived from an adult cell.

The controversy has been triggered by a letter in *Science* describing Dolly as an "anecdote not a result", referring to the fact that it was the only successful birth out of several hundred attempts. The letter was written by Norton Zinder, a bacterial geneticist at Rockefeller University in New York, and Vittorio Sgaramella, a scientist at the University of Calabria in Italy (*Science* 279, 635; 1998).

The two researchers criticize what they claim was the poor characterization of the donor cells. They restate the possibility that Dolly could have originated from a stem cell



Wilmut: "small possibility" that cloning was the result of inadvertently using a fetal cell.

— as noted in Wilmut's paper — and perhaps even from a circulating fetal cell, as the donor ewe was pregnant at the time the cells were obtained. They argue that the lamb's origins would have been better clarified by a genetic fingerprinting comparison with the donor ewe.

In fact, cells used in the cloning experiments were taken from a ewe that died three years before Dolly was born, and Wilmut's team did not directly compare Dolly's DNA with that of her clone. Ron James, managing director of PPL Therapeutics, which spon-

sored Wilmut's work, admits that in retrospect it would have been prudent to have taken more precautions.

But he explains that the mammary gland cells used had not been better characterized because they were prepared for other studies aimed at increasing expression of milk proteins, and had not been created with nuclear transfer experiments in mind.

This explanation cuts little ice with critics such as Richard Gardner, an embryologist at the University of Oxford. "I don't believe this was an 'after-tea' experiment," he says. "It should have been better documented." Gardner describes the absence of a direct comparison between Dolly and her clone as a "staggering omission", leaving open the possibility of a mix-up in the cells used. Others who have criticized the relative lack of evidence include the Nobel laureate Walter Gilbert of Harvard University.

Wilmut has argued that a mix-up of cell cultures was unlikely, given that Dolly is a Finn Dorset ewe, and that no other Finn Dorset cells were being cultured in the laboratory. But he admitted last week that, although the prospect is highly unlikely, there is a small possibility that Dolly could have originated from a fetal cell, and that he may repeat the experiments.

In response to the criticisms, PPL has asked an unnamed US university to verify that Dolly's mitochondrial DNA came from the Scottish Blackface breed of sheep used as a source of unfertilized eggs, and an unnamed British team to compare Dolly's DNA with that of stored samples of the udder cells.

A definitive answer should come from a comparison of Dolly's DNA with tissues from the donor ewe. Wilmut says frozen samples were kept, and are being studied by an independent group. "Other samples will also be available to interested parties if they wish," he says.

Many cloning scientists have come to Wilmut's defence. James Robl, a scientist at Texas Transgenic Nuclear Transfer Calves, describes Zinder and Sgaramella's criticisms as "nitpicking". He says they are "out of touch" with cloning research. Robl argues that, far from being an "anecdote", one birth from 277 attempts is a "very respectable result in this field where there are many steps, each of which take their toll on efficiency".

Robl dismisses as highly improbable the possibility that Dolly might have originated from a fetal cell, given that the sheep placenta contains many layers of tissues separating the fetus from the mother's blood.

Ken White, a cloning scientist at Utah

French agency faces fresh controversy

[PARIS] An inquiry by France's ministry of national education, research and technology into the activities of a laboratory of INSERM, the national biomedical research agency (see *Nature* 391, 519; 1998), is likely to take a new turn following a request from two scientists that their names be removed from a key paper based on work at the laboratory that is being considered for publication.

In the letter dated 17 February to the US journal *Proceedings of the National Academy of Sciences* (PNAS), the scientists from the INSERM Laboratory of Nutrition, Lipoprotein Metabolism and Atherosclerosis at the University of Rennes 1 ask that their names be removed from a paper entitled "Leptin acutely regulates postprandial lipemic responses in ob/ob and db/db mice" and from any revised versions.

"The main point is that we disagree with the use and misinterpretation of our data by Mr B. Bihain [the director of the laboratory]," wrote the two researchers. "Moreover, despite the fact that we are ... authors, we did not take part in the [editing] of the paper and we did not sign any document for this submission."

In an even blunter letter sent in January

to Jacqueline Godet, director of life sciences at the ministry and a participant in the inquiry, the authors say that they are unable to accept being co-authors of "incorrect" results. "Now that these falsified data may appear in an international publication, casting doubts on our scientific integrity, we have taken the liberty of asking you what we should do," they write, enclosing a series of annexes detailing the conduct of research within the laboratory.

Nick Cozzarelli, the chief editor of PNAS, says the paper is at the first stage of its review process, which is conducted by a member of the academy, in this case Richard Havel from the School of Medicine at the University of California, San Francisco. The paper has so far received favourable reviews from referees, although it has not yet been submitted to the editorial board.

"Of course, we would never leave on a paper the names of individuals who wish to have them removed," says Cozzarelli. He adds that, without wishing to prejudge a paper that has not yet been received by the editorial board, "given the history of the case, it is highly likely that the board would re-evaluate the suitability of the paper if it gets to us".

Declan Butler & Olivier de Gandt

State University, says: "Retrospectively you can say that we can't go back and check, but it is always easy to look back. I don't think that Ian Wilmut needed to do more." He adds: "I wouldn't be so quick to run up red flags saying that no-one has been able to reproduce the work. It's too soon. If by 2000 no-one has duplicated it, I'll start being concerned. But am I concerned in 1998? Not at all."

Zinder and Sgaramella argued that their scepticism had been provoked by the fact that no-one has reproduced the results. But David Wells, a scientist at the Ruakura agriculture research centre in New Zealand, who has reproduced Wilmut's work on cloning from fetal fibroblasts, and is engaged in cloning adult bovine cells, describes the controversy as "premature".

Wells says it is unfair to expect rapid confirmation of the results, given the time and money needed for such experiments, the long gestation periods of farm animals and the seasonality of their reproduction.

Bob Moore, who has recently retired as

deputy director of the Babraham Agricultural Institute near Cambridge, says he is not surprised that other scientists have not yet repeated Wilmut's finding, given the difficulties of obtaining success. "I think people are having difficulties reproducing the results, but you only have to look at [the low efficiency of] Wilmut results to predict that they should be having difficulties."

Moore describes the controversy as a "storm in a teacup". He points out that Dolly is part of a growing body of science that has witnessed a logical progression from the production of clones from embryos, and differentiated fetal fibroblasts, to the routine generation of embryos from adult somatic cells, even though, apart from Dolly, none have gone to term.

Jean-Paul Renard, a scientist at the French national agricultural agency INRA, who is attempting to clone cattle from adult cells, is even more confident. "Our work leads us to think that Dolly is a reality, and will be reproduced," he says.

James argues that, contrary to popular belief, few groups are racing to reproduce the Dolly experiment. For commercial purposes, cloning from fetal fibroblasts is more efficient, while their genetic modification is relatively straightforward, he says. "It delivers what we need — a genetically modified animal." Wilmut's group has so far not attempted to repeat the experiment simply because "Dolly is not commercially useful," James argues.

"You are talking at least \$1 million" to repeat the Dolly experiment, says Robl. He is also pursuing cloning from fetal fibroblasts as the most immediately promising commercial option, while carrying out research to develop more efficient systems for cloning from adult cells.

Zinder remains unconvinced by such arguments, asserting that the extensive media and political reaction to cloning has obscured the fact that "the emperor has no clothes". He says Wilmut should have been required to repeat the result before his paper was published.

Declan Butler

Genome panel defends researchers' — and families' — interests

[WASHINGTON] The latest recommendations from the international organization that coordinates the work of genome researchers are meeting criticism from those seeking to defend patients' privacy. The organization wants scientists to retain the ability to gather more information about DNA samples from subjects' medical records, and researchers to be able to overrule subjects in some cases by informing family members about their findings.

The ethics committee of the Human Genome Organisation (HUGO) also recommends that research samples obtained with consent and stored "may be used for other research" if subjects are informed at the time of the original donation and do not object — and provided that information identifying the subject is either replaced with a code allowing the individual to be traced, or stripped from the samples.

But HUGO's statement expresses strong reservations about the irreversible 'stripping' of information. Even if it is appropriate in certain circumstances, it should be done with caution, "since it may preclude valuable uses of the samples and validation of results".

The retention of codes allowing subjects to be traced would mean that a subject might be contacted years after participating in a research trial by researchers pursuing another study. Some privacy advocates argue that this should happen only in rare, extenuating circumstances. Virtually all oppose it as a general policy.

Many scientists feel strongly that irreversible loss of access to subject information seriously damages a sample's

value. For instance, stored tumour samples might be found to contain a particular oncogene, but its prognostic value would be impossible to discover if subjects or their records could not be traced.

Bartha Maria Knoppers, a law professor at the University of Montreal and chair of the HUGO ethics committee, says the statement seeks to define an ethical middle ground between allowing researchers unfettered access to samples and attached clinical information, and requiring specific informed consent for each use of a sample.

The statement was drafted by scientists, ethicists and lawyers from ten countries. Its publication comes as governments, international organizations and societies around the world are working to define ethical standards for the use of DNA samples that neither hobble research nor trample on patients' rights.

In the United States, researchers can at present gain access to tissue samples gathered for other research, or through clinical practice, by persuading local ethics boards that they are essential to a study, and that their use will not harm subjects.

The ethics committee studied some 75 statements by private, national and international groups before coming to its conclusions. It will present them to HUGO members at a meeting in Turin, Italy, late next month. The HUGO statement is not legally binding but "is meant to inspire national approaches", says Knoppers.

HUGO's caution about irreversibly stripping identifying information from samples, and its recommendation that researchers should be able to use samples

collected for one purpose for other studies, is welcomed by Mark Sobel, chief of the molecular pathology section at the US National Cancer Institute and president-elect of the Association for Molecular Pathology. The recommendations are "wonderful," he says. "They would be consistent with what most research communities and the pathology community would want."

But others say the statement bends too far towards researchers' interests. Retaining links to enable a researcher to obtain more information raises "real concerns about the invasion of privacy" that have not been given "adequate acknowledgement or weight" by the HUGO committee, says Ellen Wright Clayton, associate professor of paediatrics and law at Vanderbilt University in Nashville, Tennessee.

Clayton was lead author of a more conservative 1995 statement by a working group convened by the US National Institutes of Health and the Centers for Disease Control and Prevention.

The HUGO statement also recommends that researchers should be allowed to overrule a subject's wishes that blood relatives should not be informed of findings, if the findings point to "a high risk of having or transmitting a serious disorder and prevention or treatment is available".

"Genetic information is both personal and familial," says Knoppers. "We can't just [exclude] the real and morally valid interests of family members." But not all researchers agree. "I have trouble with overriding an individual's need for privacy except in an immediately life-threatening situation," says Sobel.

Meredith Wadman

Medical researchers set up UK body to raise public profile

[LONDON] An organization representing the professional interests of academic medicine is to be set up in Britain in an effort to raise the public profile of the biomedical sciences.

The establishment of the Academy of Medical Sciences was announced last week by a coordinating committee whose members include representatives of the Royal Society, the royal medical colleges, and university medical schools.

This group, which has been meeting for the past two years under the chairmanship of Peter Lachmann, the biological secretary of the Royal Society, will act as the academy's interim council until it formally begins its activities in October. The academy will initially have 350 founder fellows, but the eventual aim is to increase the fellowship to 1,000 over the next ten years.

According to proposals drawn up by the coordinating committee, the academy will seek to promote the practical application and public understanding of medical research. It also plans to advise the government and other public bodies on issues related to medical sciences that pervade policy decisions.

One important role will be to bridge the gap between basic research in the biomedical sciences and clinical practice. "We aim to create an integrated body which can provide expert advice on issues of national interest," Christopher Edwards, principal of Imperial College School of Medicine and a member of the coordinating committee, said at a press conference last week.

Previous attempts to establish a professional body for the medical sciences have met with opposition from some parts of both the Royal Society and the medical colleges, on the grounds that it would duplicate some of their own functions. But, according to Edwards, recent issues emerging from scientific advances, such as cloning and bovine spongiform encephalopathy (BSE), have highlighted a need for a professional body representing the interests of medical research.

Members of the coordinating committee say that issues likely to be addressed by the academy include the effect of biotechnology on the medical sciences and the ethical aspects of cloning. It may set up a subcommittee on ethical issues, bringing together a wide range of experts and perhaps carrying out joint studies with other organizations.

Welcoming the creation of the new academy, Sir Aaron Klug, president of the Royal Society, said that "the lack of a unified voice for academic medicine has been apparent for some time particularly on those issues of science, science policy and science support that embrace all the sciences". **Asako Saegusa**

Japan backs carbon project in shift to applied research

[TOKYO] Japan's Ministry of International Trade and Industry (MITI) is actively backing an ambitious large-scale national research and development project designed to produce 'highly functional carbon and related materials'.

The move reflects a shift in the focus of MITI's support for research and development from basic towards more applied research, as part of its efforts to help restructure and stimulate Japan's economy by contributing to the growth of new industries.

The five-year project into the possible application of carbon materials — including fullerenes, carbon nanotubes, carbon nitride and diamond-like carbon — is one of several industry-orientated research projects being launched this year by MITI's Agency of Industrial Science and Technology (AIST).

MITI officials say there has been a policy decision to encourage the establishment of new competitive industries in response to the current poor economic conditions, following the cabinet approval last May of plans to reform the Japanese economy.

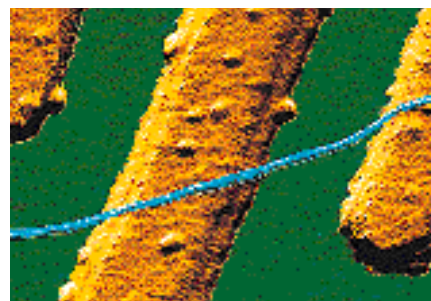
Harutoshi Yamada of AIST says that building new industries and technologies is a priority for the government. The relative lack of venture-capital-based businesses, combined with a reluctance by university professors to establish their own companies, means the ministry must actively encourage promising technologies, he says. "In Japan, everything cannot be left to the market."

The government is backing this strategy. Despite an austere budget for the new fiscal year that starts in April, leaving overall government expenditure virtually flat, MITI won a relatively large increase of 4.3 per cent in its research and development budget (see *Nature* 391, 111; 1998).

Over the past few years the powerful ministry, internationally renowned for its active industrial policy and encouragement of new technologies, has increasingly focused its attention on nurturing basic research and turning its 15 national research institutes into international "centres of excellence".

But ministry officials acknowledge that the focus has switched to creating new technologies and industries. The new Frontier Carbon Technology Project, an illustration of this shift, will be based at MITI's National Institute of Materials and Chemical Research (NIMC) in Tsukuba, northwest of Tokyo.

The project will receive ¥1.4 billion (US\$12 million) each year over a five-year period. Researchers will work on synthesis and materialization technologies for highly functional carbon and related materials. Yoshinori Koga, a senior researcher at the



Narrow research: the world's smallest electrical wire is just a dozen or so carbon atoms wide.

institute, says this could lead to the development of new devices, semiconductors and materials for use in flat panel displays.

The project has received wide support from academics and industry in Japan, which has relatively strong research on nanotubes and fullerenes. Yohji Achiba, a leading Japanese chemist and researcher into fullerenes at Tokyo Metropolitan University, argues that the time has come to make a national effort to promote such research, as it is essential for the country's future to take the financial risks involved in supporting research that might not deliver the results anticipated.

One company to welcome the project is the electronics manufacturer NEC, a world leader in research into fullerenes and carbon nanotubes. "We would like to expand our research activities through the project," says Hisatsune Watanabe, vice-president of NEC's research and development group.

He says the project offers the opportunity to create new devices to control photons, such as scanning near-field optical microscopes, as well as the chance to work with companies from different fields.

Not all supporters of the carbon project are optimistic that it will lead to the creation of new industries. Although it might be politically important for MITI to be seen to be expanding industrial activity in Japan, argues one senior industry expert, the creation of these industries "is not so easy", as applications are still a long way off.

The project will be funded through the New Energy and Industrial Technology Development Organization and will link researchers at NIMC with industry and universities in a focused group under a new 'concentrated research and development system'.

Other new MITI research projects being launched this year include a large-scale genome programme aimed at developing technologies for decoding and using genetic information, and a superconductor application technology programme. **Richard Nathan**

Germany seeks 'non-modified' food label

[MUNICH] As European Union member states continue to argue about how genetically engineered foods should be labelled, Germany and Austria are each struggling with the problem of how non-genetically engineered foods should be defined and labelled.

Farmers and some food manufacturers would like to be able to label their products as being free of genetically modified organisms (GMOs), but are afraid to do so without clear rules about permissible levels of contamination with foreign DNA.

Given the increasing number of GMOs in the environment, they say, even the most careful producers cannot guarantee that their products will be free of contamination by these organisms' DNA. The producers want to be able to sell their products as GMO-free despite this unavoidable contamination.

A head-on confrontation emerged in Bavaria, Germany, last week between a coalition of environmental pressure groups and religious groups, and the Bavarian government, led by the conservative Christian Social Union (CSU). The critics want a system of labelling that would allow a small amount of contamination. But the government is advocating a system requiring evidence of no contamination at all before a label is approved.

The coalition says that such a requirement would undermine the concept of labelling because no food producer would be able to risk making such a claim. To undermine labelling in this way, it says, would undermine consumer choice. The coalition accuses the Bavarian government of acting in the interests of major food producers, which oppose the concept of labelling according to the presence or absence of GMOs.

But the Bavarian government says that allowing any level of contamination in a food labelled as GMO-free would deceive consumers. Last December, the government put forward a bill to regulate labelling of GMO-free foods, which it sees as "a model for regula-

"GMO-Free Unless Accidentally Contaminated, These Things Happen From Time To Time, You Know, Even In Bavaria."



tion at the federal and European Union (EU) levels which are at the moment missing".

The coalition, Gentechnikfrei aus Bayern ('produced without using genetic engineering techniques from Bavaria'), has collected enough signatures to allow it to begin formally to initiate a referendum on whether the electorate would accept low-level 'accidental' contamination in food labelled GMO-free.

The coalition proposes that food products qualifying for the Gentechnikfrei aus Bayern label should follow strict production rules. For example, any animal whose meat is used must not have been fed on genetically engineered soya beans or cereals. Samples must also be regularly analysed for purity.

A similar initiative in the north German state of Lower Saxony, Gentechnikfrei aus Niedersachsen, has been received more sympathetically by local politicians. But the state government agrees with Bavaria that the local origin of the product — Bavaria or Lower Saxony — should not be mentioned on the label. It wants to move as soon as possible to federal, or even EU, rules.

Lower Saxony is one of several social democrat *Länder* (states) that have asked the Bundesrat, Germany's parliamentary upper house, to draw up federal legislation on the

labelling of non-genetically engineered foods. They are proposing that a low level of contamination should be allowed, without a defined threshold, on the grounds that appropriate tests are not available. 'Low level' would be defined on a case-by-case basis.

Horst Seehofer, the federal health minister, has promised to draw up a list of criteria for labelling food as GMO-free by the beginning of next month. So far, however, Seehofer, who is a member of the Bavaria-based CSU, has not indicated what position he is likely to take on contamination levels.

Austria is the only EU country close to legislating on how GMO-free products should be defined and labelled. Last December a coalition called AGL, incorporating supermarkets, food manufacturers, organic farmers and environmental groups, produced criteria for controlling the production of GMO-free foods, and set a level of permissible contamination of protein encoded by foreign genes at 0.1 per cent of product weight. AGL spokesman Florian Faber says that this level was chosen fairly arbitrarily.

Some Austrian food producers have already started to use the label to which the AGL certification procedures entitles them. But this month an advisory panel to the Austrian food regulatory office suggested stricter rules for controlling production.

Unlike AGL, the panel would not grant a GMO-free label to cheeses made with chymosin (rennin) derived from the stomachs of calves that had been reared on milk from cows fed with genetically engineered food. The panel also suggests that the extent of allowable 'unavoidable contamination' should be decided case by case according to the availability of suitable analytical tests.

Contamination is mainly likely to occur through imported animal feeds or storage in containers that previously held genetically manipulated products. Currently, as in Germany, virtually no genetically engineered crops are grown in Austria, making contamination through cross-pollination unlikely.

Most scientists agree that it is virtually impossible to rule out completely the risk of DNA contamination, given the widespread presence of GMOs that have been deliberately released and the difficulty of accurately detecting low levels of contamination.

Biotechnology companies are rushing to develop sensitive tests for potentially contaminating genes and the European Commission is supporting research on the topic.

There is an irony in the apparent inversion of values about labelling GMO-free foods, says Faber. "Organic farmers and consumer protection groups are demanding the right to have contamination, while big food companies are saying there should be none," he says.

Alison Abbott & Burkhardt Roeper

ANZAAS looks to grassroots for relaunch

[SYDNEY] The 110-year-old Australian and New Zealand Association for the Advancement of Science (ANZAAS) is seeking to re-establish itself following the election of a new council last week. The former council members resigned en masse two months ago when they failed to secure sufficient votes for their recommendation to dissolve the organization (see *Nature* 391, 4; 1998).

The association's new chair is Paul Adam, a biologist at the University of New South Wales and head of one of the three

state divisions that had opposed dissolution.

The association's aims — to promote communication, interaction, public awareness and the curiosity of children — remain "as relevant and important as ever to the future of science in Australia", says Adam. Initially, ANZAAS will refocus its activities at the local and regional levels, with an emphasis on serving young scientists and science teachers, rather than mounting national congresses. Such congresses failed to attract significant numbers in recent years.

Peter Pockley

Ecologists seek flexible protection rules

[WASHINGTON] The Clinton administration last week formalized its controversial 'no surprises' policy under which landowners who take steps to preserve natural habitat are guaranteed that they will not incur additional expense or obligations under species protection laws. Scientists, meanwhile, are pressurizing lawmakers to correct what they see as flaws in the policy.

The Endangered Species Act (ESA) allows landowners to 'take', or harm, some protected animals or plants in exchange for developing Habitat Conservation Plans (HCPs) that improve the species' overall chance of survival. About 225 such plans are already approved and 200 more are planned.

The interior secretary, Bruce Babbitt, has pushed HCPs as a way of involving private property owners in species protection, as half of the more than 1,000 species considered at risk of extinction are exclusively on private land. In 1994 he began promising landowners that there would be "no surprises" once an HCP was approved; the plans would be binding, in some cases for as long as 100 years.

Last week the Department of Interior codified 'no surprises' as a formal policy. But the policy has been criticized (see *Nature* 386, 530; 1997) because it does not allow for changing circumstances or new scientific



The grey wolf: prominent on 'endangered' list.

information about an endangered species.

Last month, 17 leading ecologists restated those and other worries in a letter to four senators who have introduced legislation to reauthorize the ESA. The authors propose "a few crucial amendments to make the [bill] more scientifically credible". They ask for assurances on stable funding for landowner incentive programmes that may require government spending, and for HCPs to be adjusted if they are not working.

They also call for the removal of two provisions from the bill: the requirement for independent scientific review of all listing decisions "even though many such decisions generate little or no scientific controversy", and the requirement for a detailed economic

analysis of recovery measures.

The letter's main authors, Gary Meffe of the University of Florida and Stuart Pimm of the University of Tennessee, were joined by the ecologists Edward O. Wilson of Harvard, Paul Ehrlich of Stanford, Thomas Eisner of Cornell, Peter Raven, and two former presidents of the Ecological Society of America, Ronald Pulliam of the University of Georgia and Gordon Orians of the University of Washington.

Several policy initiatives announced with the 'no surprises' regulation suggest that the Interior Department is taking the criticisms seriously. The initiatives call for "expanded use of adaptive management for all HCPs", the establishment of clear biological goals for the plans, and improved scientific monitoring. The department is also to consider limiting the duration of some HCPs. Draft guidelines for these initiatives are expected to be published within the next two weeks.

If the new legislation is to succeed, the trick will be to make the 'no surprises' policy flexible while still assuring landowners that they will not be surprised. Debate about ESA reauthorization is at present deadlocked, with both property rights advocates and environmentalists unhappy with the bill introduced by Dirk Kempthorne (Republican, Idaho) and three other senators.

Tony Reichhardt

Boost to biodiversity research 'would strengthen US economy'

[WASHINGTON] Biodiversity research in the United States should be boosted because it will lead to a healthier environment which will, in turn, strengthen the nation's economy, an influential panel of scientists says.

In a report requested last year by President Bill Clinton – and to be delivered next month – the President's Council of Advisors on Science and Technology (PCAST) says that spending on biodiversity research should grow from \$460 million a year to \$660 million over the next three years. The report was prepared for PCAST by a panel chaired by Peter Raven, the director of the Missouri Botanical Garden in St Louis.

"I think that the logic of [the report] is compelling," says Murray Gell-Mann of the Santa Fe Institute in New Mexico, a member of PCAST and of Raven's panel. He adds that the report's suggested "synergy" between the environment and the economy "will resonate with people, including members of Congress".

In the recent past, the discipline has been attacked by conservatives in Congress, who see the gathering of information on biodiversity as a threat to landowners' property rights.

As a result of such attacks, for example, the National Biological Survey was disbanded by the Congress in 1995, and the Senate has declined to ratify the international Convention on Biological Diversity, the agreement signed at the Earth Summit in Rio de Janeiro in 1992, which has already been ratified by 161 other countries.

The report emphasizes the relationships between biodiversity conservation, a healthy environment and a strong economy. It also calls for the development of better computer networks and databases for information about plant and animal species.

Five major funding proposals are suggested by the panel for consideration by the dozen or so federal agencies involved in

biodiversity research, over the next three years. They are:

- an increase from \$74 million to \$130 million a year in funding for taxonomists to discover and describe new species;
- an increase from \$300 million to \$355 million for research and monitoring of ecosystems;
- \$24 million in new money for social science research, chiefly at the National Science Foundation, to improve estimates of the economic value of sound environmental management;
- "a minimum of \$40 million a year" to develop a 'next generation' National Biological Information Infrastructure on which information about species can be stored and accessed;
- spending on environmental education to rise from \$72 million to \$87 million, mostly to train 10,000 schoolteachers a year about environmental science.

The recommendations are largely based on discussions with government scientists about their resource needs, and these

scientists are pleased about the outcome. "For us, this is a breath of fresh air," says Michael Ruggiero, a senior ecologist at the Biological Resources Division of the United States Geological Survey, which has absorbed the National Biological Survey.

Consideration by PCAST, Ruggiero says, "may be the highest level at which this issue has ever been addressed".

If previous PCAST studies are any guide, the Raven panel is likely to have significant influence on next year's budget requests from the agencies involved. But changing Congress's approach will be a larger challenge.

"There's a lack of general appreciation of what scientists are coming to understand about connections between the loss of biodiversity and the things that people care about," concedes Jane Lubchenco of Oregon State University, a panel member. The report, she adds, is "just one stage" in changing that appreciation.

Colin Macilwain

US urged to explore accelerator options

[WASHINGTON] The United States is not ready to build any more new particle accelerators, two separate panels of scientists have told the government. But it should press on with the conceptual design of a new linear accelerator, as well as researching options for other machines that might be built early in the next century.

While these ideas are being developed, the panels say, the top priorities in high-energy physics should be US participation in the Large Hadron Collider (LHC) project at CERN, the European Laboratory for Particle Physics in Geneva, Switzerland, and securing increased funding for particle physicists at universities.

The High Energy Physics Advisory Panel (HEPAP), which advises the Department of Energy on its \$700 million-a-year high-energy physics programme, last week endorsed one of the reports, prepared by a panel chaired by Frederick Gilman of Carnegie-Mellon University in Pittsburgh.

The panel says that the Stanford Linear

Accelerator Center in California should create a conceptual design for an electron-positron linear collider "in close collaboration". But it says emphatically: "This is not a recommendation to proceed with construction."

At the same time, a National Research Council panel chaired by Bruce Winstein of the University of Chicago issued a substantially similar set of recommendations. Like HEPAP, it calls for the development of a conceptual design for a linear accelerator with an collision energy of up to 1.5 TeV.

Recommending that research should be "vigorously pursued" into technologies for future muon and very large hadron colliders, the Winstein panel says this effort "should focus on a reduction of cost through the use of advanced technologies". The report does not express a view on which of these concepts is most promising.

Energy department officials welcome the modest consensus struck by the reports.

"Everybody agrees we need to bring the design forward before approving construction" of a new linear collider, says Bob Diebold, HEPAP's executive secretary.

Winstein says his panel's 160-page report was supposed to make the case for particle physics to Congress and a wider public.

"We feel it is our obligation to speak more clearly and more frequently to the public," Winstein says.

US plans to participate in the LHC project have already received some unwanted attention from the Congress. Money for the LHC was included in a list of "questionable funding increases" issued by Bob Livingston (Republican, Louisiana), the chair of the Appropriations Committee in the House of Representatives.

Officials had expected trouble for the project in Congress, but they believe that Livingston's prime motive is to use the international LHC agreement to obtain leverage with the administration, not to end US participation in the project. **Colin Macilwain**

UK nuclear physicists set to lose their privileged funding status

[LONDON] The United Kingdom's dwindling band of nuclear physicists are to lose their protected status within the science budget under proposals set out yesterday (25 February) by the Engineering and Physical Sciences Research Council (EPSRC).

For the past five years, nuclear physicists have been allocated about £5 million (US\$8 million) of research funding each year, distributed by the nuclear physics panel of the EPSRC. The budget was guaranteed partly as a quid pro quo for the government's decision to make nuclear physics the responsibility of the EPSRC instead of the Particle Physics and Astronomy Research Council when the United Kingdom's science funding councils were reorganized in 1993.

From next year, however, this small community of 60 researchers will have to compete for funds alongside other physics disciplines, with the applications assessed by a general panel of EPSRC physicists. The research council believes it is time that nuclear physics was judged on its merits. A spokesman adds that the changes will allow the discipline to "grow or recede according to the quality of proposals".

Nuclear physicists have mixed reactions to the proposed change. Few are concerned about the quality of research proposals — and thus their likely success in broader competition — describing the UK's nuclear physics as "among the best in the world". But some fear that the change in funding arrangements could divide a close-knit



Unstable future? Nuclear physics research will have to compete alongside other disciplines.

community of researchers, and deprive the field of its long-term focus.

John Durell, for example, head of the nuclear structure group at the University of Manchester, says that ring-fenced funds allow different research groups to meet and decide which areas to focus on before applying for funds. They also give research teams time to plan their research, and make advance bookings of international nuclear structure facilities.

Indeed, the Institute of Physics is concerned that a general EPSRC panel of physics researchers may not be the best way of assessing nuclear physics research proposals. "The institute has received comments from the community that existing panels do not have the knowledge to deal with the current range of proposals," says a spokesman. "The inclusion of nuclear physics in the panel will exacerbate the problem."

Despite the current enthusiasm for particle physics, nuclear physics — which deals with the interactions between protons, neutrons and electrons — still has much to offer, say researchers. On the theory side, for example, fundamental questions about nuclear structure remain, such as why atoms change shape when neutrons are added or taken away.

Another area, nuclear isomers, is concerned with excited nuclear states that live for long periods. "If we can release this energy in a controlled way, we may have a possible energy store," says Phil Walker, professor of physics at the University of Surrey, and leader of the university's nuclear physics group.

The relatively small size of the nuclear physics community partly reflects the discipline's status as a 'poor relation' to particle physics. But it is also due to the United Kingdom's lack of a low-energy particle accelerator — needed to probe the structure of the nucleus — since the Nuclear Structure Facility at Daresbury in Cheshire closed in 1993 (see *Nature* 362, 278; 1993).

The head of one nuclear physics group says that the lack of an indigenous facility remains more of a threat to UK nuclear physics than any change in funding arrangements. "We need to find a way to build a new British facility," he says.

"It's embarrassing in that we are sponging facilities overseas, but not offering anything in return. It is as if we were a Third World country." **Ehsan Masood**

Iraq crisis spurs new bioweapons moves

David Dickson

Lack of provision in the Biological Weapons Convention for monitoring and verifying compliance has long been a worry. But negotiations for such provision have been given new momentum by recent events in Iraq.

[LONDON] The recent crisis over Iraq seems to have boosted efforts to reach international agreement on ways to monitor the spread of biological weapons.

Those seeking to curb the spread of such weapons have long criticized the lack of any provision in the Biological Weapons Convention (BWC) of 1972 for monitoring and verifying the compliance of signatory states. Negotiations have been taking place since 1995 on a protocol to the BWC designed to achieve just this.

But until now, efforts to implement a verification regime had been held back by those countries — in particular the United States — that have claimed verification to be impractical, and a potential threat to both national security and the commercial confidentiality of their pharmaceutical and biotechnology industries.

Late last month, however, in his State of the Union address, President Bill Clinton spoke of his keenness to see negotiations on the protocol speeded up. In particular, he indicated for the first time that the United States is prepared to accept the concept of on-site inspection of both declared and undeclared sites.

In reaching this decision, Clinton has opened the way for detailed discussion between Western allies of the type of biological weapons verification regime that they would like to see introduced. "We are not yet at the beginning of the end [of negotiations]; but in a sense, this is the end of the beginning," says one diplomat.

No direct link has been made to concern about reports of stocks of anthrax and other agents held by Saddam Hussein. But there is a widespread impression that the critical situation in Iraq has been an important catalyst for the US decision by placing the issue of biological weapons at the top of the political agenda.

Until recently, for example, it was being said in Washington that strong opposition to any form of on-site monitoring from the pharmaceutical and biotechnology industries, supported by their allies in Congress, was likely to block any further progress.

But Clinton's statement, and subsequent explanatory information released by the White House, indicate that, at least as far as the administration is concerned, the voice of

international prudence has received the support of the 'higher echelons' in a tense inter-agency negotiation.

Several major obstacles remain to agreement on a protocol. The pharmaceutical industry, for example, is said to have already indicated its uneasiness with the Clinton administration's position.

Perhaps even harder to persuade, some observers think, will be the defence establishment, including some of those working on biological weapons defence in the national weapons laboratories, who remain deeply concerned at the prospect of opening themselves up to international inspection.

And the United States still has to reach a common position with its Western allies, let alone other major powers. Indeed, even European states have yet to reach agreement on what a protocol should contain (although the chances of this being achieved relatively soon are said to be high, given that the United Kingdom — long an ardent advocate of a strict verification regime — is currently the president of the European Union).

One example of the gap between the United States and the United Kingdom is the question of the decision-making procedure

that should be adopted for so-called 'challenge' inspections. Britain is keen to adopt a 'red light' process as already exists for the Chemical Weapons Convention. Using this process, if a country is suspected of violating the convention, an inspection could automatically take place unless blocked by a certain number of countries.

The Clinton administration, however, has said it prefers an alternative, 'green light' approach, requiring a minimum number of countries to vote for an inspection. This position would be more difficult politically to achieve — and is thought by some to have been introduced as a concession to opponents of on-site inspections.

Further details of the US position are likely to be clarified at a high-level meeting next month between the administration and the pharmaceutical industry. In parallel, US diplomats will be meeting other Western countries in an attempt to forge a common position they can present to the Ad Hoc Group that has been working since last summer on the draft text of a protocol.

Although Clinton said in his speech that he would like to see negotiations on the protocol completed by the end of the year, this is widely seen as optimistic. Indeed, some are concerned that, having suddenly decided to stop dragging its feet, the United States now runs a different risk, of upsetting other countries by attempts to impose its own formula for a protocol on the negotiations.

But even those with reservations about the details of the US position say the chances of agreeing a protocol detailing an international verification regime on biological weapons have markedly improved. If the Iraq crisis turns out to have had any silver lining, it may be that.

David Dickson

Individuals face weapons work indictment

[LONDON] Academics concerned with arms control are floating preliminary proposals for a treaty that would make it an international crime for any individual to help a state to develop biological weapons. The treaty would apply, for example, to a researcher coming to the state from a foreign country.

One of the loopholes of the current anti-biological weapons regime is that legal sanctions are left to individual signatory states. This can create a major problem if such a state — Iraq was forced to ratify after the Gulf War in 1991 — proves reluctant to act against those suspected of

being engaged in biological weapons work.

The proposal for a new international treaty, which would enable other states to indict such individuals if found on their territory, has been drafted by a group of researchers at Harvard University in the United States and the University of Sussex in the United Kingdom.

"Our idea is to try to find a means to hold individuals responsible," says Matthew Meselson, professor of natural sciences at Harvard and a long-time campaigner against chemical and biological weapons. "These could be heads of state, or they could be the so-called

'evil professors' willing to help another country develop biological weapons."

Meselson points out that there are several significant precedents for making such activity an international crime, such as the already agreed outlawing of aircraft hijacking and theft of nuclear materials.

Many details would have to be negotiated among potential signatories. A key point, for example, is deciding which court would have ultimate jurisdiction.

But a draft of a possible convention has already been drawn up and is scheduled for eventual submission to the United Nations General Assembly.

D.D.

Governments end AIDS placebo trials in developing countries

[WASHINGTON] The United Nations, the US government and France's AIDS agency called a halt last week to the use of placebos in overseas trials seeking to prevent transmission of HIV from pregnant mothers to their infants.

The move follows a statement from the US Centers for Disease Control and Prevention that conclusive interim data from a Thai study showed that a \$50 course of zidovudine (AZT) in late pregnancy cut perinatal HIV transmission by half.

The governments said this conclusion justified the immediate halt of the use of placebos, or their replacement with short-course AZT regimens, in trials in other developing countries.

Until now, they added, placebo use "was the only way in which it could be clearly and quickly established whether a shorter AZT regimen was safe and more effective than no treatment at all" in developing countries.

Critics have attacked the use of placebos because long-course AZT treatment was shown to be effective in 1994, and is widely used in industrialized countries.

"It is inexcusable that they ever did these

experiments in the first place," says Sidney Wolfe of the Washington-based lobby group Public Citizen.

Satellite setback for Japan's space agency

[TOKYO] Japan's space programme received another setback last weekend when its H-II rocket failed to put into orbit one of the world's largest telecommunications satellites. The second burning time of fuel in the rocket's second-stage engine was shorter than anticipated, and the 2.2-tonne, ¥46.2 billion (US\$370 million) Communications and Broadcasting Engineering Test Satellite (COMETS) failed to reach geostationary orbit.

The failure is another blow to Japan's National Space and Development Agency. Last July, its environmental monitoring satellite ADEOS — the Advanced Earth Observation Satellite — ran out of power and stopped working, and budget cuts brought an abrupt end to the agency's plans to develop the unmanned space shuttle HOPE.

US study puts spotlight on radon cancer risk

[WASHINGTON] Indoor radon contributes to about 12 per cent of lung cancer deaths each year in the United States, according to a

report from the National Research Council (NRC). And most of these deaths occur among smokers. "Radon — particularly in combination with smoking — poses an important public health risk, and it should be recognized as such," says Jonathan Samet of the Johns Hopkins University in Baltimore, Maryland, who chaired the study.

About one-third of the estimated 15,400 to 21,800 radon-related lung cancer deaths each year could be prevented if radon levels in homes were reduced to the recommended Environmental Protection Agency standard of four picocuries per litre of air. Some six per cent of US homes have levels higher than that, according to the report.

NASA's chief scientist says it's 'time to move on'

[WASHINGTON] The top scientist at the US space agency NASA announced his resignation last week, after five years heading the agency's Office of Space Science. Wesley Huntress, a planetary scientist who began his career at the Jet Propulsion Laboratory in California, expects to depart in the autumn. His successor has yet to be named.

Huntress's tenure has been marked by a shift from large, expensive space science missions to smaller and less expensive ones. Although space science budgets remain tight, they are projected to rise relative to other

sectors of NASA, and several initiatives are planned. With space science in comparatively good shape, wrote Huntress in a letter to colleagues last week, "it is just simply time for me to move on".

Fusion project could go back to drawing-board

[MUNICH] The council of the 'next-step' fusion reactor project, ITER, seems to be conceding for the first time that there is a need to consider alternative reactor designs. The concession follows falling financial and political support for the project (see *Nature* 391, 620; 1998), which is a collaboration between Europe, Japan, the United States and Russia.

At a meeting last week, the council agreed to set up a working group to consider options for restricting the aims of ITER as currently designed, enabling it to be built more cheaply. But the group will also investigate the more radical solution of considering possible alternatives to ITER's tokamak design.

French 'water memory' scientist loses libel case

[PARIS] Jacques Benveniste, the French researcher who claimed in 1988 to have shown that extreme dilutions of antibody

solutions could retain their biological activity, this week lost libel suits against the Nobel prizewinners, Georges Charpak and François Jacob, and Claude Hennis, a researcher at the School of Industrial Physics and Chemistry in Paris.

Benveniste had brought the charges following statements by the scientists in articles in the newspaper *Le Monde* suggesting that his research may have been fraudulent (see *Nature* 389, 427; 1997). The *tribunal de grande instance* of Paris did not rule on the substance of the charges, but threw them out because of a procedural error — that Benveniste should have filed a penal suit and not a civil one.

One result is that Benveniste cannot bring new charges based on the articles. "I find myself in the position where I am unable to defend my honour," he says. He intends to demand that INSERM, the national biomedical agency, should launch an inquiry to clear his name. He has also challenged the scientists to repeat their statements in public.

Senators seek to double NIH budget

[WASHINGTON] Four US senators are calling on the Senate Budget Committee to allocate sufficient funds to double the budget of the National Institutes of Health (NIH) over the coming five years. In a letter to Pete

Domenici (Republican, New Mexico), the chairman of the Senate Budget Committee, and Frank Lautenberg (Democrat, New Jersey), the committee's senior Democrat, the four senators also urge that the committee allocate enough money in its 1999 budget resolution for NIH to be granted a \$2 billion, 14.7 per cent increase. Senators Bill Frist (Republican, Tennessee), Connie Mack (Republican, Florida), Dianne Feinstein (Democrat, California) and Edward Kennedy (Democrat, Massachusetts), are now soliciting signatures from other senators.

Japan offers Mars trip – but in name only

[TOKYO] More than 50,000 people have applied to have their names inscribed on a nameplate that will be affixed to Japan's first-ever spacecraft to Mars.

The Institute of Space and Astronautical Science has guaranteed that every name submitted on or before 28 February will be inscribed on the plate.

The names will be photographed, reduced to microscopic size and then etched onto the plate, which measures 2.5 cm square and 0.5 mm thick. The Mars orbiter, PLANET-B, is due for launch in October 1999.

Its primary mission will be to observe the atmosphere of Mars.

Flow of novel genes for R&D continues

Sir — Your article “Curtain falls on gene sequencing deal” (*Nature* 391, 621; 1998) contains some comments misattributed to a SmithKline Beecham (SB) spokesperson that may have created an inaccurate impression of the status and progress of the SB/Human Genome Sciences (HGS) collaboration.

In particular, we do not feel that we have “exhaustively milked” the SB/HGS database, as quoted in the article. Indeed, we firmly believe that many novel genes of therapeutic interest remain to be discovered in this resource. However, SB is focusing its drug discovery efforts on targets that we have already identified, as a matter of strategic focus. We want to be clear that we are actively pursuing many targets obtained from the SmithKline Beecham/Human Genome Sciences database, and we continue to regard the collaboration as highly productive.

We hope that these comments clarify the status of SmithKline Beecham’s collaboration with Human Genome Sciences, and resolve any ambiguity that may have arisen.

David C. U’Prichard

(Chairman, Research and Development)
SmithKline Beecham plc,
One New Horizons Court,
Brentford, Middlesex TW8 9EP, UK

Sir — Your article contains statements that are misleading and incorrectly attributed. For example, SmithKline Beecham (SB) does not claim that the company has “exhaustively milked the database”; the situation is that, having initiated work on more than 3,000 novel human genes provided by Human Genome Sciences (HGS), its capacity to pursue additional, high quality leads for drug development is temporarily saturated. The flow of novel genes for research and development continues unabated to HGS’s other partners, namely Schering Plough, Takeda, Merck KGaA and Synthelabo.

The article wrongly attributes to SB the sentiment that “the stringent terms imposed by HGS at the time” on researchers wishing to gain access to the database have become “increasingly unacceptable”. The terms of the material transfer agreements (MTAs) that govern access to the data and materials were mutually agreed by all parties to the agreement — including SB — and have been accepted by more than 150 academic institutions. HGS has already established MTAs regarding more than 1,200 individual human genes and purified proteins with academic centres worldwide.

I would also disagree that HGS triggered a “vigorous debate over whether human genome should be mapped in the private or

public domain”. HGS is not — nor has it been — a genome mapping company as are Genset, Millennium, Sequanna, Myriad, deCode, Darwin and others. HGS isolates and sequences expressed messenger RNAs; we do not map them.

Your article also quotes self-congratulation from public database providers on the completeness and adequacy of the publicly available data. But, if the public databases are adequate, why do pharmaceutical companies continue to pay large sums for access to private databases? The pharmaceutical giant Novartis, for example, signed an agreement for access to a private database in January this year.

Finally, you state that HGS is a less prominent player in the sequencing business “following the split from... The Institute of Genome Research (TIGR)”. In fact, less than 5% of the HGS human gene sequencing data was generated through our collaboration with TIGR, and we continue to be a world leader in the isolation and characterization by partial and full-length sequence analysis of human cDNAs.

William A. Haseltine

(Chairman and Chief Executive Officer)
Human Genome Sciences Inc.,
9410 Key West Avenue,
Rockville, Maryland 20850, USA
e-mail: william_haseltine@hgsi.com

New Zealand science

Sir — Your article dealing with the outcome of New Zealand’s 1992 structural reforms of science administration was of considerable interest (*Nature* 391, 426–427; 1998). It contained several points, however, that are likely to contribute to a misreading of the New Zealand situation.

For example, in contrast to the impression given by statements in your article (and the subsequent correction to one of the figures quoted), 349 new science positions have been established since the Crown Research Institutes (CRIs) were set up, whereas the number of additional support positions has risen by only 83 over this period.

Your article quotes criticism of the need for CRIs to make a profit. But such profits are reinvested in research and CRI facilities. Last year, the institutes generated after-tax profit of NZ\$29.8 million (US\$17.6 million), all of which was reinvested.

The competitive nature of project bidding under the Public Good Science Fund (PGSF) demands that CRIs are

efficient managers. The establishment of profit targets is a proven administrative tool for encouraging good performance. It should be remembered that CRIs compete for PGSF funding with tertiary institutions that are not required to operate in the commercial marketplace.

Universities are having difficulty in funding equipment through their core funding streams, as well as from the ‘contestable’ science market. But CRIs have expanded their net asset base from NZ\$139.6 million to NZ\$246.5 million. Is there any reason why academic institutions cannot manage their own performance in a similar manner?

The article did not mention a critical difficulty with the reforms, namely that universities and other tertiary educational institutes can now access the PGSF funds that the CRIs need to survive. CRIs by contrast cannot access tertiary sector research funds.

It is this one-sided nature of the science ‘market’ that has led to an increase in the number of players and suggestions that there are too many people chasing too little money. Science career paths have

not so far been significantly destabilized by this factor, although the possibility that they will is the main cause of concern that many scientists have about the reforms.

CRIs are tightly focused on science and research closely related to the needs of the New Zealand agricultural, forestry, fishing and manufacturing industries, as well as to international trade and environmental requirements of central government and local authorities. It is this responsiveness to community need that is a significant element in demonstrating to those who control government budget allocations that investment in research and science offers a better return to society than expenditure in many of the other areas that compete for funds from the public purse. Without such responsiveness, it is unlikely that science funding would have been maintained, let alone increased.

P. M. Hargreaves

(President)
New Zealand Association
of Crown Research Institutes (Inc.),
PO Box 1578, Wellington,
New Zealand

A test for the insurance industry

Life insurers claim they stand to lose unless allowed access to applicants' genetic information, but consumers insist such data should remain out of bounds. Could both sides be right?

Robert J. Pokorski

The private insurance industry has been criticized over the past decade for its insistence on using genetic information about applicants when drawing up premiums for life insurance policies. Politicians and consumer groups worry that use of genetic information may lead to a class of people unable to afford insurance. They point out that certain genetic factors deemed unfavourable may not on their own determine a person's life expectancy. And they feel that the pendulum of risk classification has swung too far in favour of low-risk policy-holders to the detriment of high-risk individuals and society in general.

Insurers, on the other hand, argue that genetic data are no different from other information routinely requested when someone applies for cover. They worry about the potential rise in applications by people who are aware of information that affects their likelihood of making an early claim but who choose not to inform their insurance company — a practice known as 'adverse selection' or 'anti-selection'. And access to genetic information has benefits, they say: a negative test result will allow a person with a family history of disease to obtain insurance at standard rates.

Here I address the conflict between actuarial and moral fairness in 'individually underwritten' policies — those that require applicants to provide personal information. Initially I defend the right of actuaries to consider genetic information. I then look at the merits of the other side of the argument, emphasizing the issue of moral fairness and challenging policy-makers to say whether people have a right to life insurance as a matter of social justice. Finally, I offer a proposal to resolve the impasse by suggesting a new type of life insurance policy in which insurance companies — under certain conditions — will not require applicants to disclose genetic test results.

Alzheimer's disease

Alzheimer's disease (AD) is increasingly relevant in the context of long-term care insurance and genetic testing. Several genes associated with this late-onset condition have recently been discovered, and genetic tests for diagnosing the disease or predicting its future occurrence are already commercially available or being developed. The Alzheimer's Association¹ maintains that "the presence of a gene [for AD] should not be a basis for underwriting insurance premiums for health care, long-term care or life insurance,

nor should it be used to infringe on any individual's access to care and services". It also recommends anonymous testing so that results would be "invisible to an individual's medical records". Similarly, a National Study Group² claims that genetic tests would be used by "the growing long-term care insurance industry to avoid insuring higher-risk individuals". These positions have prompted insurers to defend the use of genetic tests on economic³ and actuarial⁴ principles.

These policy statements reflect legitimate concern that genetic testing might limit access to long-term care (LTC) insurance. But the suggested remedy — preventing insurers from using genetic tests for AD — is in effect a call for an entitlement to buy LTC insurance regardless of one's likelihood of submitting a claim. The problem with this proposal is that private insurance is voluntary. Consumers enter and leave the market at will, switch insurers, opt for more or less comprehensive cover, and make purchases according to personal finances and expectations of claiming.

If insurers adopted the proposal, both total LTC claims and the percentage of LTC claims resulting from AD would increase, even though the incidence of AD in the general population remained the same. A positive test result might persuade more individuals to buy LTC cover because they would be told that the test indicates a high risk of AD. Claims would increase because more high-risk people would enter the insurance pool; they would buy products providing large benefits, long durations of cover and short waiting periods; and the time between buying a policy and making a claim would decrease because some people might delay purchase until AD has already been diagnosed by neuropsychological tests that detect disease in the preclinical and early clinical stages.

Unexpectedly short durations between purchase and claim are especially costly: insurers will pay out far more in benefits than they receive in premiums. If an LTC policy was issued to a 65-year-old who was already cognitively impaired (but not to the extent that he or she needed to go into a nursing home), the value of the expected claim could be more than seven times the value of expected premiums paid, and roughly six times greater if the policy was purchased at the age of 75 (ref. 5).

The market-place would compound the problem. The accuracy of genetic tests predicting AD is likely to increase. More people will be tested, will gain confidence about

their risk status, and will base purchasing decisions on their perceived risk. And policy-holders who test negative for AD will feel they have a low risk of disease, so some might decide not to buy LTC insurance.

There are financial considerations too. Insurance purchases conform to basic economic rules, with lower or higher prices encouraging or discouraging participation. The market for LTC insurance is already limited by its expense: just 6% of senior citizens have LTC insurance⁶, and cover would be affordable for only 19% of US citizens aged 55–79 (ref. 7). Overall, higher premiums resulting from increasing AD claims and fewer low-risk policy-holders in the pool will make LTC cover less attractive⁸.

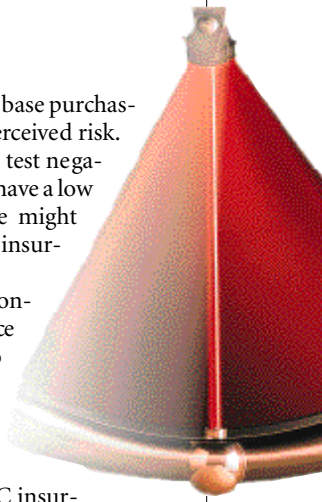
Who is entitled?

Most people believe insurance companies should offer life cover to all applicants regardless of their health or future risk, including those with a terminal illness. This emphasizes the limitations of a private system in which people want access to affordable cover while retaining the right to delay purchase until the insured event becomes likely.

Some companies already sell such a product. 'Simplified issue, graded benefit, whole life' (permanent) insurance requires no doctor's statement, medical examination or blood tests. Most people can buy it providing they have not been disabled in the past year or treated for 'internal' cancer in the previous 5 years, and are not hospitalized, terminally ill, confined to a nursing home or HIV-positive. Benefits are graded during the early years of the policy for death from natural causes, with the full amount paid if death occurs after the graded period or if the policy-holder dies accidentally at any time.

The table illustrates simplified issue cover in California. A 45-year-old man could buy \$450,000 of this insurance for a premium of \$43 per \$1,000 of cover, compared with \$21 per \$1,000 for traditional whole-life policies that are individually underwritten and pay the full benefit on natural or accidental death. The difference is due to underwriting, with premiums lower for policies with medical questions, examinations or blood tests.

Simplified issue policies do not allow unconditional access to life insurance because they are too expensive. The problem



is that companies cannot force people to buy cover; guaranteed life insurance protection would be cheaper if people were legally required to buy cover by a certain age. But in a voluntary system, an entitlement to purchase life insurance could be accomplished only through higher premiums. The question is this: "Would you be willing to forgo underwriting and pay more than twice as much for the right to buy insurance without regard to health or risk?" The answer is clear in the limited appeal of existing simplified issue products, as compared with robust demand for individually underwritten policies. People would like cover regardless of their claim risk, but they are not willing to pay the required premiums.

Fairness

The relative importance of actuarial rating and mutual pooling is central to the debate about risk classification for individually underwritten products. Actuarial rating refers to cost allocation, with policies priced so that premiums are proportional to expected benefits. Supporters of such 'actuarial fairness' stress that policy-holders with the same or similar risk should be charged the same amount. Mutual pooling, on the other hand, involves collective sharing of risks of unexpected contingencies (such as premature death or disability) and redistribution of pool funds to policy-holders who experience the insured event. Those favouring such a system speak of social inclusion, a right to insurance cover and the fact that premiums should be the same for participants of a given age regardless of individual risk.

If insurers were not allowed to use genetic information in underwriting, there would be increased variability in mortality risk among policy-holders (requiring higher premiums to cover greater risk uncertainty), a shift in insurance cover to people at higher risk, possible adverse selection by individuals who decide to purchase a particular type of cover (or a larger amount) after learning they have unfavourable genetic characteristics, and higher average mortality among policy-holders.

Actuaries can predict the probability and magnitude of these events, and their impact on premiums, consumer buying practices and the viability of the insurance plan. But are these events and their effects desirable for society? The balance between actuarial rating and mutual pooling is a matter of social policy, with the caveat that certain requirements must be met if a plan is to be

self-sustaining in the private sector.

Occasionally insurers have offered life insurance when society deemed cover essential even though this required a loosening of actuarial arguments. In Australia, life insurance is compulsory in certain superannuation (retirement) plans. The Institute of Actuaries of Australia states that insurers "may not seek medical information and members of the plan do not have to disclose [medical or genetic] test results". In the United Kingdom, some bank loans cannot be obtained without mortgage insurance. The Association of British Insurers says: "Any genetic tests which are to the detriment of applicants for life insurance up to a total of £100,000 will be ignored when the insurance is linked to a new mortgage for a home." In France there is an agreement between the Federation of French Insurers and the Ministries of Trade and Health to provide "loan security policies [mortgage insurance]" to HIV-infected individuals. The assumption here is that no one buys a retirement plan or a home simply to obtain insurance cover.

Social justice

What of society's needs? People are warned of the financial consequences of illness, disability, premature death and outliving their assets, yet are not offered the desired range of affordable protection. The philosopher John Rawls asserted that all citizens have fundamental rights as a matter of social justice. Policy-makers must decide if insurance is one of these rights, and, if so, which types of cover should be provided. They must then find ways of satisfying these obligations.

Striking a balance

To resolve the impasse, the insurance industry should work with policy-makers, the medical establishment and consumer groups to develop a life insurance policy that does not require applicants to reveal previous genetic test results. This proposal echoes previous suggestions⁹⁻¹¹ that a reasonable amount of life insurance should be available to everyone.

What features might one expect of such a policy? First, genetic tests would have no bearing on insurability, providing there was no clinical evidence of disease. Applicants would not be asked about previous genetic tests, and genetic test results in medical records would not be considered. Second, applications would be individually underwritten, with premiums based on factors

The insurance industry should work with policy-makers, the medical establishment and consumer groups to form a life insurance policy that does not require applicants to reveal previous genetic test results.

such as age and the presence of disease processes — including those with a genetic component — that had already developed. Applicants carrying a disease gene would be sold a life insurance policy at standard rates if in good health, and charged a premium commensurate with their risk if a disorder had been diagnosed that affected their claim risk. Third, this product would be sold with the stated intention of allowing consumers to purchase cover without needing to divulge genetic information. Genetic test results that might be learned in conjunction with this product would not be transferable to applications for other types of cover. For other 'traditional' forms of life insurance, companies would be allowed to use results of previous genetic tests in underwriting. Fourth, the maximum cover available from all insurers in this plan would be determined by consensus and cover value indexed to an economic indicator.

Implementation should depend on the policy's acceptance by policy-makers, the medical establishment and consumer advocates. If companies market insurance that guarantees access to affordable life cover regardless of genetic risk, the issue then is whether insurers will receive broad support for their request to use genetic test results when underwriting other individual life insurance policies. □

Robert J. Pokorski is at the Cologne Life Reinsurance Company, 30 Oak Street, Stamford, Connecticut 06905, USA.

e-mail: pokorski@colognere.com

The opinions in this paper are those of the author and not necessarily those of the organizations with which he is associated.

1. Alzheimer's Association. News release, 30 June 1995.
2. Post, S. G. et al. *J. Am. Med. Assoc.* **277**, 832-836 (1997).
3. Pokorski, R. *Nature* **376**, 13-14 (1995).
4. Pokorski, R. *Am. J. Hum. Genet.* **60**, 205-216 (1997).
5. Murtaugh, C. M., Kemoer, P. & Spillman, B. C. *Inquiry* **32**, 271-284 (1995).
6. Moses, S. A. *Natl Underwriter* 5 August 1996.
7. Weiner, J. J. *Am. Med. Assoc.* **271**, 1525-1529 (1994).
8. Norton, E. C. & Newhouse, J. P. *J. Am. Med. Assoc.* **271**, 1520-1524 (1994).
9. National Center for Human Genome Research, NIH publ. no. 93-3686 (1993).
10. Harper, P. S. *Lancet* **341**, 224-227 (1993).
11. Moseley, R. et al. Report of the Human Genome Insurance Project, April 1995.

Simplified issue, graded benefit, whole life insurance available in California

Age (years)	Companies selling cover	Maximum issue amount (\$)	Cost of insurance (\$ per \$1,000)
25	1	50,000	26
35	4	325,000	32
45	6	450,000	43
55	9	510,000	62
65	9	510,000	101

Clouds, contrails and climate

John H. Seinfeld

Cirrus clouds warm the Earth. Their formation is incompletely understood, but recent studies hint that they can evolve from jet contrails. This is an unsuspected, but possibly important, effect of human activity on climate — offering a new way to study the processes involved.

More than half of the solar radiation reflected by the Earth comes from clouds. So how might increasing levels of anthropogenic aerosols — sulphates, organics and trace gases such as nitric acid — affect the formation of cirrus clouds over the planet and thereby alter the radiation balance?

Cirrus clouds form in the upper troposphere when supercooled water droplets or ice crystals nucleate around aerosol particles, mostly aqueous H_2SO_4 droplets. When the temperature is below 0°C , air that is saturated with respect to water is supersaturated with respect to ice, so if an ice crystal is produced in a cloud that consists mainly of supercooled water, it will grow very much more rapidly than its supercooled-water neighbours. As ice crystals grow more and more rapidly, they may reduce the water vapour pressure below saturation relative to supercooled water, leading to the rapid transfer of water from droplets to crystals. This process concentrates the available water in a small number of large ice crystals.

But human emissions of aerosols may increase droplet concentration in clouds by increasing the number of particles that can serve as cloud condensation nuclei. That would increase scattering and therefore cloud-top reflection, and thus reduce the amount of energy absorbed by the Earth. Unfortunately, the magnitude of this effect on the global radiation budget is highly uncertain, estimated¹ to be between 0 and -1.5 W m^{-2} (for comparison, the effect of anthropogenic greenhouse gases is about $+2.4 \text{ W m}^{-2}$). The uncertainty largely stems from our poor understanding of how aerosols affect the nucleation of different types of cloud.

Low-level stratus clouds usually cool the planet, because they reflect more solar radiation back to space than they absorb rising infrared radiation. By contrast, cirrus clouds — ice-particle clouds that lie between a few thousand metres up and the tropopause, at 10 to 20 km — exert an overall heating effect, because their reflectivity at solar wavelengths is low, and they absorb more rising infrared radiation than they emit to space from their very cold cloud tops. So if anthropogenic gases and aerosols change the properties and

lifetimes of clouds, they could have a positive or negative radiative effect, depending on the extent to which low- and high-level clouds are affected. Worse, we cannot yet predict how increasing levels of trace gases and particles will influence the properties, extent and lifetime of any type of cloud. But a laboratory to study how this happens is provided by jet aircraft.

Jet contrails may be viewed as artificial cirrus. Aircraft emit oxidized sulphur, soot, metal particles and nitrogen oxides, and contrails form as the hot, humid exhaust gases mix with the cold, dry air. Soot and sulphuric acid particles in the exhaust are activated into aqueous droplets that subsequently freeze. Sulphur and soot emissions from aircraft are small compared with global sources, but in the altitude range of emissions (about 10 to 12 km), aircraft sources are estimated to be comparable to natural and anthropogenic sources². Thus, aircraft emissions could substantially affect cirrus formation. Because of their small area, individual contrails cannot

appreciably affect the Earth's radiation balance, but if large areas of persistent cirrus are produced from contrails, then jet exhaust may have a greater effect on climate than is currently believed.

Several recent field programmes have tried to assess the atmospheric impact of aircraft exhaust^{3–6}. Results of NASA's SUCCESS project ("Subsonic aircraft: Contrail and cloud effects special study"), conducted in April–May 1996, will appear in several forthcoming issues of *Geophysical Research Letters*⁶. Satellite observations of contrails during SUCCESS show that persistent contrails can indeed evolve into extensive cirrus clouds that otherwise would not have formed⁷. This means that a considerable amount of ice-supersaturated air exists in the upper troposphere that is capable of producing cloud in the presence of nucleating aerosols. How often contrails spawn extensive cirrus clouds, and how large and long-lived those clouds are, are still open questions.

The nitrogen oxides emitted by jet engines may be partly responsible for promoting cirrus formation. Based on nitric acid measurements in the upper troposphere, Laaksonen *et al.*⁸ have calculated the effect of nitric acid, HNO_3 , on activation of supercooled droplets in cirrus. They find that in the presence of the observed amounts of HNO_3 , 1.5 to 4 times as many H_2SO_4 and $(\text{NH}_4)_2\text{SO}_4$ aerosol particles should experience activation as in the absence of HNO_3 . The climatic implications of these findings depend on the frequency of cirrus formation by the supercooled droplet activation mechanism, and

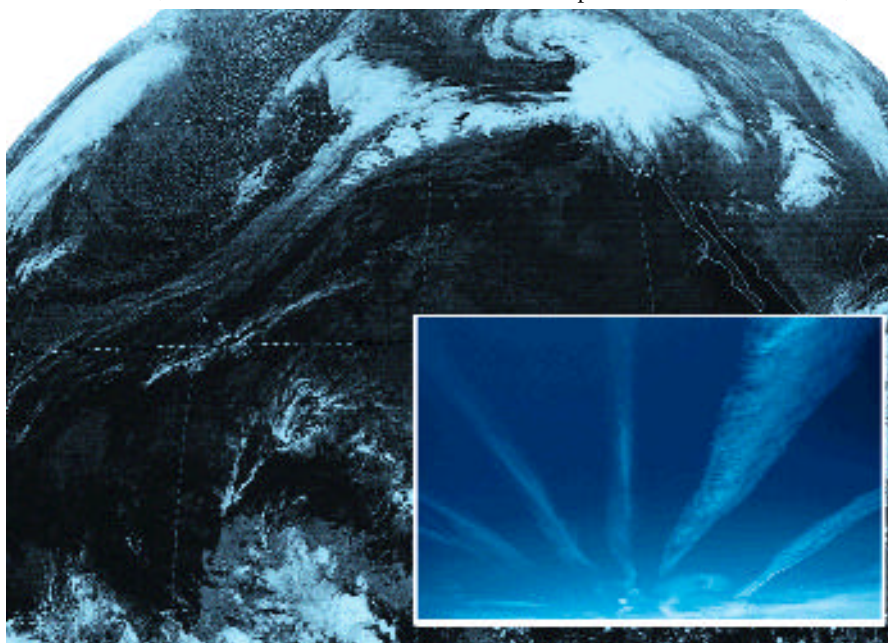


Figure 1 Clouds over the north Pacific Ocean. The GOES West satellite measures cloud cover by detecting $10\text{-}\mu\text{m}$ infrared radiation. Cirrus clouds are abundantly evident in this image, along the US west coast and over Hawaii, for example. Anthropogenic aerosols, including jet contrails (inset), may affect the formation of cirrus, perhaps appreciably affecting the Earth's climate. Indeed, it has now been shown that jet contrails can evolve into extended cirrus. But the mechanisms of cirrus cloud formation, and its participation in upper-atmosphere chemistry, are still highly uncertain.

UNIV. WASHINGTON

TONY STONE

that is not known. Nonetheless, the trace gas HNO_3 may be important in cirrus formation — and the nitrogen oxides emitted by aircraft at cruise altitudes are all converted to HNO_3 .

Incidentally, nitric acid and cirrus clouds could be linked in another way. Global, three-dimensional models of tropospheric chemistry tend to predict higher HNO_3 levels in the upper troposphere than those observed. It may be that cirrus ice crystals scavenge nitric acid and then settle to lower levels in the atmosphere, where the HNO_3 is liberated.

Might other aerosols be important? Sulphate is the predominant aerosol in the upper troposphere, but as much as 25% by mass is other species, such as crustal compounds, metals, soot, ammonium and nitrate^{9–13}. Such particles can act as cirrus cloud condensation nuclei if they are as efficient as H_2SO_4 and $(\text{NH}_4)_2\text{SO}_4$ in nucleating ice. This is not known, but *in situ* measurements made in cold (–35 to –60 °C) cirrus clouds over Germany revealed that most aerosol particles are below 0.3 μm in diameter¹⁴, and all the 'other aerosol' substances will probably be present as small particles. Even if such particles nucleate crystals inefficiently, they must play an important role in cirrus formation because of their large number.

So we have learned that the upper tropo-

spheric aerosol consists of more than just sulphuric acid and water, but measurements are scarce. We are also learning, slowly, about the mechanisms of formation of cirrus clouds. How changes in tropospheric aerosol concentration, including aircraft emissions, will affect all types of cloud is a vital question in assessing the overall climatic impact of anthropogenic emissions. □

John H. Seinfeld is in the Division of Engineering and Applied Science and the Department of Chemical Engineering, California Institute of Technology, Mail Code 104-44, Pasadena, California 91125, USA.

e-mail: seinfeld@cco.caltech.edu

1. Houghton, J. T. *et al.* (eds) Intergovernmental Panel on Climate Change (IPCC) *Climate Change 1994* (Cambridge Univ. Press, 1995).
2. NASA (National Aeronautics and Space Administration) *Scientific Assessment of the Atmospheric Effects of Stratospheric Aircraft* NASA Ref. Publ. 1381 (1995).
3. Busen, R. & Schumann, U. *Geophys. Res. Lett.* **22**, 1357–1360 (1995).
4. Schumann, U. *et al.* *J. Geophys. Res.* **101**, 6853–6869 (1996).
5. Schumann, U. *Ann. Geophys.* **12**, 365–384 (1994).
6. Toon, O. B. *et al.* *Geophys. Res. Lett.* (in the press).
7. Minnis, P. *et al.* *Geophys. Res. Lett.* (in the press).
8. Laaksonen, A. *et al.* *Geophys. Res. Lett.* **24**, 3009–3012 (1997).
9. Heintzenberg, J. *et al.* *Atmos. Res.* **41**, 81–88 (1996).
10. Talbot, R. *et al.* *Geophys. Res. Lett.* (in the press).
11. Coggiola, M. *et al.* *Geophys. Res. Lett.* (in the press).
12. Chen, Y. *et al.* *Geophys. Res. Lett.* (in the press).
13. Twohy, C. H. & Gandrud, B. W. *Geophys. Res. Lett.* (in the press).
14. Ström, J. *et al.* *J. Atmos. Sci.* **54**, 2542–2553 (1997).

Brown and Morgan¹ look at the herbivore end of the equation, by feeding squirrels with seeds spiked with toxins. And in the latest *Journal of Ecology*, Heil and co-workers² examine the proportion of photosynthetic productivity that a plant needs to expend to support a population of mercenary ants that will protect it from grazers.

Generalist herbivores are more successful when they adopt a mixed diet. One hypothesis for this is the complementarity of food sources — as the proportion of one component of the diet falls, its value to the consumer rises, making it worth seeking out. This may be due to the nutritional content of the food resources, or the presence of different toxins in the components of the diet. Consumption of an excess of one toxin could, for example, lead the grazer to seek plants with a different chemical constitution.

Schmidt *et al.*¹ developed a model based on these ideas, then, by experimentation in nature, tried to determine whether complementarity really does occur. They chose to work with grey and fox squirrels (Fig. 1), in mixed woodland and meadow habitats in Illinois. The squirrels were given mixed sunflower seeds impregnated with toxins (tannic and oxalic acids). Squirrels are exposed to these toxins in their normal feeding — tannins in oak acorns, for example, and oxalates in various leaf tissues and fungi. The sunflower-seed base has a constant nutritional value, allowing chemical factors alone to influence food selection, and the toxins in the seeds were at a similar concentration to those found in the squirrels' natural foods. (Control seeds were soaked in distilled water.) Using various combinations of test seeds and controls, and by examining the residue of the seeds at the end of a day's exposure to squirrels, the feeding preferences were quantified.

The authors found that squirrels neglected the oxalate-impregnated seeds more than those impregnated with tannin, showing a greater dietary sensitivity to oxalates. But when the supply of tannin was augmented (by setting out an extra tray with just the seeds containing that toxin), fewer oxalate-treated seeds were left behind at the end of

Ecology

Green policies for defence spending

Peter D. Moore

Defence expenditure is a problem that faces plants as well as politicians. Spend too little and the aggressor takes advantage; spend too much and vital

resources are wasted. Plants defend themselves from aggressive herbivores in many ways — ranging from the conventional weapons of spines and stinging hairs, to more sophisticated chemical and biological deterrents. These include toxins, or the development of mutualistic relationships with animal defenders such as ants. All of these methods cost energy, and the products of photosynthesis must be diverted from plant growth or reproduction into defence. But how can the expenditure be optimized, and what is the response of the herbivore to deterrents? Writing in *Oikos*, Schmidt,



Figure 1 Sitting targets — plants are at the mercy of grazing herbivores such as the fox squirrel, *Sciurus niger*. But Schmidt *et al.*¹ have found that, by producing certain toxins, plants can force grazers to seek alternative sources of food.



Figure 2 A *Pseudomyrmex* worker visiting the food bodies on an acacia leaf. Some plants employ the services of aggressive ants for protection against grazing herbivores. Heil *et al.*² have calculated the amount of its total productivity that a plant invests in this defence mechanism.

the day. Complementarity is thus occurring — when exposed to an excess of one toxin, the grazer moves to food sources containing alternative toxins. This means that the evolutionary development of chemical defences by plants effectively forces grazers into generalist foraging strategies. It also means that 'dietary preference' is a relative term, because it varies according to the proportions of the alternative food sources available.

Because an investment in secondary-metabolite toxins diverts the attentions of the herbivore away from the target plant, it is energy well spent. But how much does a plant need to invest in defence to ensure its own safety and yet avoid energetic profligacy? The energetic costs involved in producing secondary plant metabolites can be calculated, but many metabolites are not used exclusively in defence systems and it is difficult to measure only defence expenditure. So Heil and co-workers² have examined the question in a different strategic situation, whereby the plant uses the services of resident, aggressively territorial ants for protection against grazers. The ants attack intrusive herbivores, whether they be invertebrate or vertebrate. The costs involved in this type of defence include the provision of residence facilities for the ants and, in particular, supplying them with food.

The authors studied *Macaranga triloba*, a pioneer tree from the forests of Malaysia. Ants inhabit the hollow stems of the plant and obtain their nutrition from 'food bodies', rich in lipids and proteins (Fig. 2). These are borne on the undersurface of stipules, along the stems, and their sole purpose is to support the ants. To determine the costs of producing these food bodies, Heil *et al.* experimentally excluded ants by ringing selected stipules with a sticky resin. The difference in food-body production between stipules with ant-exclusion and those with ant-access over a selected period gives the rate at which food bodies are removed. The authors then compared the energetic costs of this supply with the overall primary production of the tree.

The results were highly variable, depending on the size of the tree, the total leaf area, the number of stipules and so on. The production of food bodies ranged from 3.8 to 10.6 mg per day, which is equivalent to 130–584 $\mu\text{M CO}_2$ per day, per plant. Because the estimated net productivity was of the order of 7–63 mM CO_2 per day, per plant, the investment in food for ants amounts to between 0.6% and 5.0% of the plant's net productivity. The proportional investment in defence was greatest in young, unbranched trees, and declined with age and complexity of branching. There is great variability (and, therefore, considerable error) involved in this preliminary estimate. Also, it takes into account only the energetic content of the bodies themselves; secretory glands

and even the construction and maintenance of the stipules could be regarded as defence investments. So, overall, this is a conservative estimate.

The fact that mutualism has evolved and been maintained between *Macaranga triloba* and ants is indicative of a cost-benefit advantage to the plant in fitness terms. Unpublished data by the authors indicate that unprotected plants are twice as likely to be grazed by herbivores as the plants with ant defences, quite apart from possible protection from shoot-borers and fungi. Secondary compounds tend to offer most protection to mature plant tissues, but it is often the young leaves that are most at risk

from grazing, and the mobility of the ants provides added protection here. Ant defence also avoids problems of plant self-injury resulting from the accumulation of toxic compounds.

The Book of Proverbs tells us that we have much to learn from the ant and its industrious ways (Ch. 6, 6–8). Perhaps defence strategists have even more to learn from the canny plant. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

1. Schmidt, K. A., Brown, J. S. & Morgan, R. A. *Oikos* **81**, 130–142 (1998).
2. Heil, M. *et al.* *J. Ecol.* **85**, 847–861 (1997).

Molecular physiology

Key clockwork component cloned

Richard W. Tsien

Brain waves and cardiac pacemakers are just two examples of repetitive electrical activities that rely on T-type calcium channels for rhythmic firing^{1,2}. Inappropriate activity of T-type channels may contribute to epileptic seizures³, and it is small wonder that an understanding of their molecular basis has been eagerly sought. Now, on page 896 of this issue, Perez-Reyes and colleagues⁴ report the cloning of an $\alpha 1$ subunit that can support classical T-type calcium-channel activity. Interestingly, this is the first case in which a voltage-gated calcium-channel subunit has been isolated with the help of sequence information from the Human Genome Project.

The isolation of a T-type channel subunit goes a long way to closing the circle on calcium-channel diversity. This began with separation of low- and high-voltage-activated calcium channels^{5,6}, quickly developing into distinctions between T-, N- and L-type

channels^{7,8} and, eventually, P-, Q- and R-type channels, after multiple forms of the $\alpha 1$ subunit were found to exist⁹. But T-type channels stand out from other voltage-gated calcium channels. They can be opened transiently by relatively weak depolarizations (positive deflections), but only if the preceding potential is sufficiently negative^{5,7}. T-type channels can, therefore, act as a key component in electrophysiological rhythm generators — somewhat like a pendulum that arcs back with precise timing in a grandfather clock.

When a neuron or cardiac pacemaker cell receives a hyperpolarizing nudge, the basal inactivation of T-type channels is removed, making them available to open and support the entry of calcium. This, in turn, depolarizes the cell, often resulting in rebound excitation in the form of a broad, 'low-threshold spike'¹. The hyperpolarizing push can be provided by inhibitory neurotransmitter, or

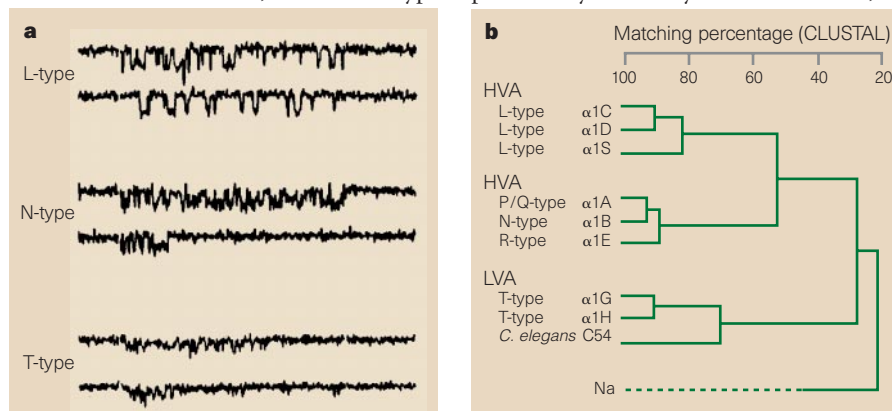


Figure 1 The molecular basis of T-type calcium channels. a, Unitary calcium currents supported by L-, N- and T-type channels in chick sensory neurons⁷. T-type channel openings were originally distinguished and named with their tiny conductance and strikingly transient time course in mind. b, Family tree of voltage-gated calcium channel $\alpha 1$ subunits, separated initially into high-voltage-activated (HVA) and low-voltage-activated (LVA) channels. The $\alpha 1G$ subunit isolated by Perez-Reyes *et al.*⁴ and the $\alpha 1H$ subunit (L. Cribbs & E. Perez-Reyes, personal communication) both support classical T-type channel activity.

by open potassium channels. T-type channels are particularly well-adapted for generating such events because they can open over a negative range of potentials — much more negative than other voltage-gated calcium channels, and considerably more negative than even voltage-gated sodium channels. Like other calcium channels, T-type channels may also work as a delivery system for calcium as a second messenger. For example, this kind of signalling is crucial for production of a sustained rise in cytosolic calcium in adrenal glomerulosa cells, leading to the manufacture and secretion of aldosterone². T-type channels may also be involved in the control of cytosolic calcium during cellular proliferation, and their expression is known to fluctuate in synchrony with the cell cycle².

Previous attempts to clone the T-type channel have encountered a number of blind alleys. Impressed by kinetic similarities between T-type currents and sodium currents, some investigators have looked for a calcium-selective channel subunit with homology to sodium-channel sequences — but to no avail. T-type calcium-channel subunits have also been predicted to form a tetramer, as in cyclic-nucleotide-gated or potassium channels. The $\alpha 1G$ subunit isolated by Perez-Reyes *et al.*⁴ has the pseudotetrameric structure and signature motifs found for other calcium-channel $\alpha 1$ subunits. These include positively charged residues in the S4 transmembrane segments (which are the key to voltage-dependent gating), and acidic residues in the P segments, which are needed to support divalent cation selectivity and permeation.

The $\alpha 1G$ subunit has low sequence homology (<30%) to other voltage-gated calcium channels (Fig. 1). This suggests an early evolutionary divergence from previously known calcium- and sodium-channel subfamilies, and also helps to explain why previous attempts at cloning failed. A homologue of $\alpha 1G$ has been found in the nematode *Caenorhabditis elegans* (C54), although it has not yet been functionally characterized. Nonetheless, Perez-Reyes *et al.* provide convincing evidence that $\alpha 1G$ is an authentic T-type channel subunit. When they expressed this subunit in *Xenopus* oocytes, the authors obtained currents with the key marks of classical T-type channels: a single-channel conductance of ~8 pS; slow deactivation (shutting-off of conductance after a sudden repolarization); and a characteristic crossing pattern of current families, evoked by a series of depolarizing voltage steps. This is not to say that $\alpha 1G$ is the only T-type calcium channel. A close relative of $\alpha 1G$, labelled $\alpha 1H$, is plentiful in heart and also supports classical T-type channel activity (L. Cribbs & E. Perez-Reyes, personal communication). Moreover, it has been suggested¹⁰ that the $\alpha 1E$ subunit may show T-type channel behaviour. So variant forms of T-type chan-

nel activity may co-exist with classical T-type channels within individual neurons^{2,3}.

The isolation of a genuine $\alpha 1$ subunit from a T-type channel opens many promising avenues for future work. The development of specific antibodies, knockout mice and channel/calcium-sensor hybrids will help us to localize T-type channels and decipher their possible role in cell proliferation. Moreover, the $\alpha 1G$ and $\alpha 1H$ subunits could help us to understand what makes a calcium channel a calcium channel. Interestingly, the critical acidic residues in the four P regions follow the pattern Glu–Glu–Asp–Asp, rather than the Glu–Glu–Glu–Glu seen in other $\alpha 1$ subunits. Following the introduction of a T-type blocker, mibefradil (Posicor)¹¹, which is effective in treating hypertension, the search will be on to find other agents that selectively inhibit T-type channels. And we can hope that agents will soon be available for T-type

channels with the potency and specificity of ω -conotoxin-GVIA for N-type channels or ω -agatoxin-IVA for P/Q-type channels. □

Richard W. Tsien is in the Department of Molecular and Cellular Physiology, Stanford University Medical Center, Stanford, California 94305-5345, USA.

e-mail: rwtsien@leland.stanford.edu

1. Jahnsen, H. & Llinás, R. *J. Physiol. (Lond.)* **349**, 227–247 (1984).
2. Tsien, R. W., Clozel, J.-P. & Nargeot, J. (eds) *Low-Voltage-Activated T-Type Calcium Channels* (Adis International, Montpellier, 1998).
3. Huguenard, J. R. & Prince, D. A. *J. Neurosci.* **12**, 3804–3817 (1992).
4. Perez-Reyes, E. *et al. Nature* **391**, 896–900 (1998).
5. Carbone, E. & Lux, H. D. *Biophys. J.* **46**, 413–418 (1984).
6. Armstrong, C. M. & Matteson, D. R. *Science* **227**, 65–67 (1985).
7. Nowycky, M. C., Fox, A. P. & Tsien, R. W. *Nature* **316**, 440–443 (1985).
8. Nilius, B., Hess, P., Lansman, J. B. & Tsien, R. W. *Nature* **316**, 443–446 (1985).
9. Birnbaumer, L. *et al. Neuron* **13**, 505–506 (1994).
10. Soong, T. W. *et al. Science* **260**, 1133–1136 (1993).
11. Clozel, J.-P. *et al. Drug Rev.* **9**, 4–17 (1991).

Optics

A time-resolved glimpse of the terahertz glory

Philip L. Marston

If you stand above a cloud, you will see bright, coloured rings around your shadow. These 'glory' rings are most commonly seen around the shadow of an aeroplane, but they can also be seen in mist from a high vantage point (Fig. 1). The investigation of these rings helped motivate C. T. R. Wilson's development of the cloud chamber. But what causes this strong backscattering of light by droplets of water? In a paper last month¹, a new method for resolving the optical scattering processes was used to confirm the importance of surface-wave contributions to the scattering.

Over 50 years ago, Van de Hulst² explained glory scattering in terms of optical waves travelling around the surface of the droplet, which radiate weakly focused toroidal wavefronts back in the direction of the observer. The theory is well developed³ but is difficult to test experimentally.

The scattering processes are complicated: there are conventional waves that are refracted and reflected at the droplet surfaces, as well as the surface waves, which 'leak' into the water, and so take short-cuts through the droplets (Fig. 2, overleaf). The leakage occurs because the tangential component of the wavevector along the short-cut matches that of the surface wave. The coupling with the surface wave occurs over a range of surface positions so that the associated rays are not spatially resolvable in experiments. Even though glory scattering has been studied for a range of systems^{4–6} (not only electromagnetic waves), direct measurements of the strength of the surface-wave/electromagnetic coupling processes are

rarely available in the optical or near-optical domain. Now Daniel Grischkowsky's group at Oklahoma State University have been able to distinguish the different scattered-wave components from their different travel times.

The authors fired electromagnetic pulses of 0.1 to 1 THz (with a spectral peak at a wavelength of about 1 mm) at a 6.35-mm-diameter alumina sphere¹. The radiation was generated by transient excitation of a dipole antenna with a short optical pulse from a laser. Another optoelectronic system, illuminated by a delayed laser pulse, allowed time-resolved reception of the scattered signal. By changing the delay time, glory rays were sep-

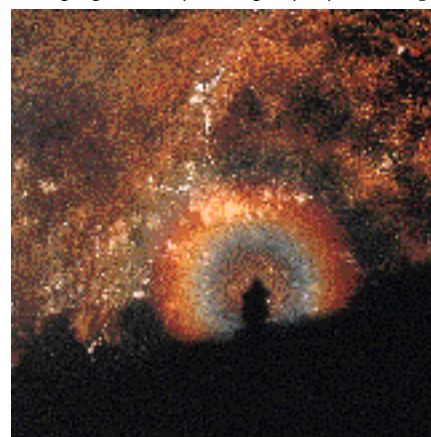


Figure 1 Optical glory: the small droplets of water in the mist backscatter the light to produce the coloured ring around the photographer's shadow. The outer portion of the ring is orange because the backscattering is wider at longer wavelengths.



100 YEARS AGO

The author of the alleged discovery in this instance is Dr. Waltemath, an astronomer of Hamburg, and his assertion is, that there is evidence of the existence of a second moon, circling about the earth, but with such low reflective power that it has usually escaped observation even when in opposition. The author has perceived the desirability of making this hypothetical moon do a useful work in celestial mechanics, by explaining that the difference between the observed and computed secular acceleration of the moon's mean motion arises from the action of the newly discovered body. ... It is, of course, quite possible that Dr. Waltemath fully believes in the existence of this object. In that case we should say, he is the only person who does; for when we ask on what kind of observation does this very accurate orbit rest, we find that the author has employed that large collection either in which persons have believed that they have seen objects of doubted value transiting the sun, whether bright or dark. He seems to have trusted to those wild and reckless assertions that are made from time to time about "ruddy fireballs" or "night suns," or other vague descriptions, and on such loose and inaccurate data he has unfolded his strange and wondrous tale.

From *Nature* 24 February 1898.

50 YEARS AGO

The Cantor Lecture before the Royal Society of Arts, delivered on January 20 by Dr. C. H. Andrewes, of the National Institute for Medical Research, is an admirable summary of the essential features of our knowledge of that world-wide scourge, the common cold; and nobody is better fitted than Dr. Andrewes to state what we know and what we do not know about it. ... If, as Dr. Andrewes showed, those who are working on this problem in Great Britain and other countries are only beginning to pick up clues to the solution of what is one of the most complex problems of virology, there is plenty of evidence that this relative ignorance will not persist for very long. All the portents, indeed, suggest that presently the common cold, like typhus, typhoid, cholera, smallpox, yellow fever and other plagues, will have yielded to human domination.

From *Nature* 28 February 1948.

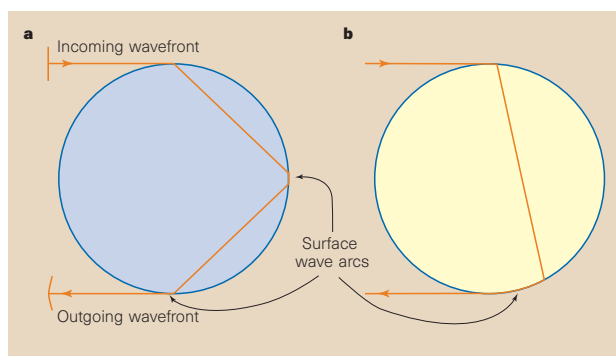


Figure 2 Ray diagrams of the principal contributions to the optical glory: a, a small, spherical drop of water (refractive index 1.33), and b, a dielectric sphere made of alumina (refractive index of 3.12). Time-resolved experiments show that the optical glory comes from surface waves that take a short-cut through the sphere.

arated from the weaker (and less interesting) direct reflections, and the spectrum of the glory could be isolated. The measured spectral contributions were consistent with theoretical models that attribute glory backscattering to the superposition of surface waves.

For the sphere investigated, the refractive index of 3.12 is much greater than the optical refractive index of water, which is close to 1.33. Consequently, the glory ray viewed had only one short-cut (Fig. 2b), which differs from visible glory rays for water droplets (Fig. 2a), but the confirmation of theory is no less valid for that.

The measurements were taken at a backscattering angle of 11° , which is offset from the most highly focused part of the scattering. If the technique can be extended to the brighter on-axis backscattering,

details of the time evolution of the surface-wave coupling process could be explored. The effects of changing the shape of the scatterer⁵⁻⁷ could also be investigated. In nature, after all, drops typically lack the symmetry of a perfect sphere. □

Philip L. Marston is in the Department of Physics, Washington State University, Pullman, Washington 99164-2814, USA.

e-mail: marston@wsu.edu

1. Cheville, R. A., McGowan, R. W. & Grischkowsky, D. *Phys. Rev. Lett.* **80**, 269–272 (1998).
2. Van de Hulst, H. C. *J. Opt. Soc. Am.* **37**, 16–22 (1947).
3. Khare, V. & Nussenzveig, H. M. *Phys. Rev. Lett.* **38**, 1279–1282 (1977).
4. Nussenzveig, H. M. *Diffraction Effects in Semiclassical Scattering* (Cambridge Univ. Press, 1992).
5. Berry, M. V. *Adv. Phys.* **25**, 1–26 (1976).
6. Marston, P. L. (ed.) *Selected Papers on Geometrical Aspects of Scattering* (SPIE Press, Bellingham, WA, 1994).
7. Marston, P. L. & Trinh, E. H. *Nature* **312**, 529–531 (1984).

Ecology

Kiss of death

What do the Denver Broncos have in common with a tiny beetle larva? A clue can be found in *Proceedings of the National Academy of Sciences* (95, 1108–1113; 1998), where Thomas Eisner and colleagues describe a cunning mechanism that is used by larvae of the phengodid beetle to evade the chemical defences of millipedes larger than themselves — and then kill them.

Millipedes deter predators by squirting a toxic fluid from defensive glands found along the length of their bodies, apart from the first five segments and the last. Nevertheless, they are routinely eaten by phengodid larvae. First the worm-like larva coils its body around the front of the millipede, embedding its mouth parts in the millipede's neck and somehow immobilizing it. Then the larva retires underground before reappearing to consume the soft inner tissues, leaving only the skeletal parts and glandular sacs uneaten.

How does the larva do this? By staging encounters between a phengodid larva (*Phengodid laticollis*) and the millipede (*Floridobolus penneri*), Eisner and colleagues have found the answer — in



piercing the millipede with its sickle-shaped mandibles, the larva delivers a lethal injection of regurgitated gastric fluid. Moreover, the millipede fails to discharge its defensive glands during the attack, and the authors found that the discarded glands of dead millipedes contain similar levels of benzoquinone toxins to their live counterparts.

So back to American football. The answer, of course, is that the Broncos and the beetle have both found a way to defeat more powerful opposition, coming away with — respectively — the superbowl and supper.

Alison Mitchell

Protein metabolism

Slow and fast dietary proteins

Gema Frühbeck

Although the physiological effects of different types of fatty acid and carbohydrate are relatively well known, the effects of protein composition on metabolic processes have been investigated less extensively. But, writing in *Proceedings of the National Academy of Sciences*, Boirie *et al.*¹ seek to close this gap in our knowledge. By analogy with dietary carbohydrates, the authors apply the concept of 'slow' and 'fast' proteins, according to the speed at which proteins are digested and amino acids are absorbed from the gut. And they conclude that slow and fast proteins differentially modulate whole-body protein deposition after a meal.

Boirie *et al.* studied the effect of two milk proteins — casein and whey protein — on postprandial whole-body protein metabolism. They combined oral and intravenous administration of intrinsically labelled and unlabelled proteins, to circumvent some of the technical problems of trying to measure amino-acid kinetics after a meal. Healthy adults were given whey protein, and the authors observed a high, rapid and transient peak in plasma amino-acid levels. This was associated with a pronounced increase in protein synthesis, but no change in protein breakdown. By contrast, after a casein meal, the plasma amino-acid profile was lower, slower and prolonged. The whole-body metabolic response consisted of a slight increase in protein synthesis, but a marked inhibition of protein breakdown. According to the speed at which amino acids appeared in the bloodstream, the authors classified whey protein as a fast protein and casein as a slow protein.

Protein turnover is a finely controlled process, in which gain or loss of body protein is adjusted to biological needs (Fig. 1). The physiological mechanisms that control the acute response to protein ingestion in humans are controversial. Several reports^{2–4} have shown an increase in amino-acid oxidation after meals (especially when subjects have a high protein intake), but with inconsistent changes in protein synthesis and breakdown. *In vivo* studies showed that breakdown is inhibited after ingestion of protein, but they failed to demonstrate that protein synthesis is stimulated substantially^{2,4–6}.

How are the rates of protein synthesis and degradation regulated by these so-called slow and fast proteins? Boirie *et al.* attribute the slow increase in plasma amino-acid levels after the casein meal to a probable delay in gastric emptying — this is consistent with the formation of casein clots in the stomach. Whey protein, by contrast, is quickly emptied from the stomach into the duodenum⁷.

The rapid influx of amino acids after ingestion of whey protein may be analogous to the dramatic stimulation of protein synthesis that is seen when the concentration of amino acids in the plasma is quickly raised by intravenous infusion⁸. Differences in gastric emptying between casein and whey protein may also be due to opioid peptides, which are released from casein during digestion. These slow down gastrointestinal motility — thus delaying gastric emptying — by interacting directly with opiate receptors in the gut⁹.

Other mechanisms may explain the differences in the inhibition of protein breakdown between whey protein and casein. For example, insulin is known to inhibit proteolysis^{2–4}. But Boirie *et al.* rule out this mechanism, because the concentrations of insulin in the plasma after zero, 40 and 300 minutes were no different between the two meals. However, the insulin-to-glucagon ratio may be more important than the absolute concentrations of either¹⁰. Casein and whey protein are both milk-derived proteins, but they have different amino-acid profiles. Although the percentage of amino acids that stimulate the secretion of insulin is the same in both proteins, the higher proportion of branched-

chain amino acids in whey protein leads it to have a synergistic effect with insulin on protein metabolism¹⁰. Moreover, the percentage of amino acids that stimulate the secretion of glucagon is substantially lower in whey protein than in casein. Hence, the catabolic effect of glucagon — which counterbalances insulin in the acute control of protein turnover — is less after a whey-protein meal. The distinct amino-acid compositions of casein and whey protein may, therefore, trigger a discriminatory stimulus on protein synthesis and breakdown due to differences in the secretion of insulin and glucagon.

The potential to increase protein synthesis and inhibit protein breakdown selectively offers a number of therapeutic possibilities for patients with wasting disorders. One of the goals in the nutritional support of patients with sepsis, extensive burns, acquired immunodeficiency syndrome, malignancy, skeletal trauma or multiple organ failure is to minimize the loss of body protein. If the accelerated net breakdown of body protein can be slowed, or even reversed, by administering specific proteins, it would represent a very attractive option for nutritional support in the critically ill patient. Moreover, the concept of slow and fast proteins may be applied if high concentrations of amino acids in the plasma are detrimental to the patient — such as in renal insufficiency or hepatic encephalopathy. Another potential application is the production of milk

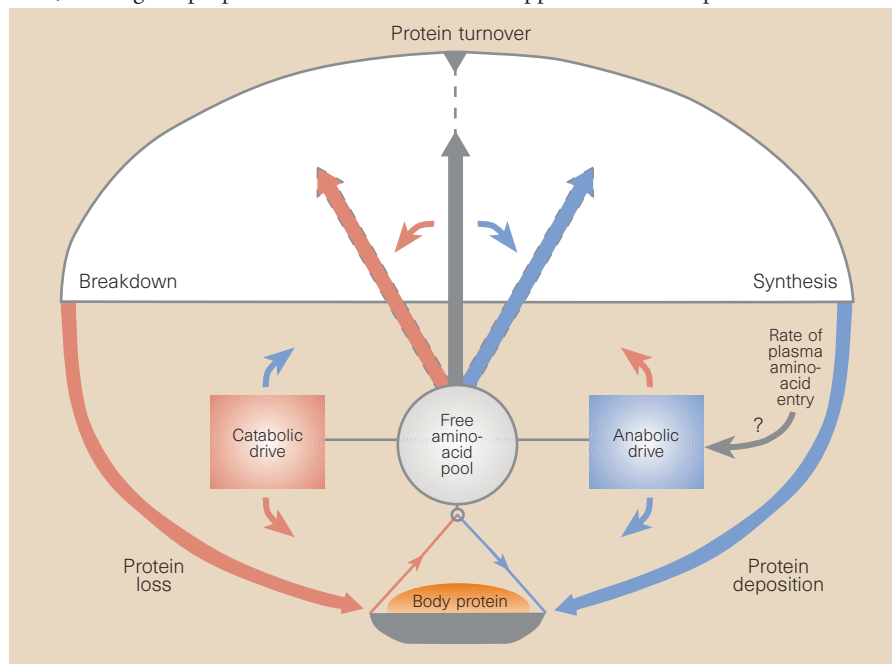


Figure 1 The two-pool model for protein turnover. This model shows how the balance between catabolic and anabolic processes exerts a homeostatic control on the free amino-acid pool, tipping the metabolic system in favour of net synthesis or breakdown. When the influence of the anabolic drive outweighs the catabolic influences, the system shifts to protein synthesis, leading to protein deposition. But breakdown and, hence, protein loss occur when catabolic factors predominate over the anabolic drive. Boirie *et al.*¹ have found that ingestion of whey protein rapidly leads to high amino-acid levels in the blood plasma, and a concomitant increase in protein synthesis. But a casein meal leads to a slower accumulation of amino acids in the plasma, and only a slight increase in protein synthesis.

formulas for pre-term infants with low body weight, to achieve the maximum growth rate and food efficiency.

Every new finding generates questions. For example, why should dietary proteins have different effects on protein turnover? Is this finding unique to whey protein and casein, or can it be applied to other dietary proteins? Are there striking differences between proteins from animal and vegetable origin? What is the site of postprandial protein deposition? Does the effect on protein metabolism differ between lean and obese people? But perhaps the first step is to determine the extent to which the effect of the amino-acid absorption rate on protein metabolism is reproducible when proteins are given under more physiological circumstances — as part of a typical, mixed meal. □

Gema Frühbeck is at the MRC Dunn Clinical Nutrition Centre, Hills Road, Cambridge CB2 2DH, UK.

e-mail: Gema.Fruhbeck@mrc-dunn.cam.ac.uk

1. Boirie, Y. *et al.* *Proc. Natl. Acad. Sci. USA* **94**, 14930–14935 (1997).
2. Garlick, P. J., McNurlan, M. A. & Ballmer, P. E. *Diabetes Care* **14**, 1189–1198 (1991).
3. Pacy, P. J., Price, G. M., Halliday, D., Quevedo, M. R. & Millward, D. J. *Clin. Sci.* **86**, 103–118 (1994).
4. Waterlow, J. C. *Annu. Rev. Nutr.* **15**, 57–92 (1995).
5. Nissen, S. & Haymond, M. W. *Am. J. Physiol.* **250**, E695–E701 (1986).
6. Young, V. R., Gucalp, C. R., Rand, W. M., Matthews, D. E. & Bier, D. M. *Hum. Nutr. Clin. Nutr.* **41**, 1–18 (1987).
7. Mahé, S. *et al.* *Am. J. Clin. Nutr.* **63**, 546–552 (1996).
8. Giordano, M., Castellino, P. & DeFronzo, R. A. *Diabetes* **45**, 393–399 (1996).
9. Daniel, H., Vohwinkel, M. & Rehner, G. *J. Nutr.* **120**, 252–257 (1990).
10. Garlick, P. J. & Reeds, P. J. in *Human Nutrition and Dietetics* (eds Garrow, J. S. & James, W. P. T.) 56–76 (Churchill Livingstone, Edinburgh, 1993).

Neurobiology

Homeostasis or synaptic plasticity?

Yves Frégnac

Almost 50 years ago, Hebb¹ proposed that during development, learning and perception, correlated activity induces long-lasting strengthening in synaptic transmission. Remarkably, however, neurobiologists inspired by Hebb's principle have sought only one side of the regulatory process — that is, synapse-specific changes in synaptic strengths. They have rarely thought about the compensatory mechanisms that regulate the total synaptic strength of a neuron (reviewed in ref. 2). But on page 892 of this issue, Turrigiano *et al.*³ describe stabilizing mechanisms that may represent a more general form of activity-dependent regulation of synaptic transmission. The changes in post-synaptic sensitivity that they have found might be seen as a demonstration of basic homeostasis, designed to return the integrative function of the cell to within a reference working range.

Five homeostatic processes related to synaptic integration are generally recognized. The simplest form regulates the efficacy of transmission around a mean synaptic gain or between two boundary values. When evaluated on a longer timescale, the fast, input-dependent regulation of synaptic transmission (recently described in cortical networks^{4,5}) results in an averaged synaptic efficacy that is roughly constant in the face of rapid changes in the probability of transmitter release⁶. Furthermore, the probability of inducing potentiation, depression or depotentiation depends on the previous stimulation history of the network ('metaplasticity' in ref. 7) and on the initial state of the synapse.

The second form of homeostasis — predicted in many models of learning — acts more globally, at the neuronal level⁸. It assumes that the sum of all the synaptic

weights of synapses impinging on the cell (or of their square values) remains constant. A third form of homeostasis limits the anatomical divergence of growing axons, corresponding to a conserved sum of the synaptic weights of all contacts made in the network by the same parent axon⁹. The fourth form of homeostasis, which is relevant to the findings of Turrigiano *et al.*³, concerns the capacity of a cell to maintain a similar output (for example, its firing level), despite strong alterations in the activity pattern of the network (reviewed in ref. 10). This regulatory process involves activity-dependent tuning of post-synaptic sensitivity, and examples so far come from the regulation of intrinsic conductances¹¹. A last form of homeostasis concerns possible changes in the site at which an impulse is initiated and the direction in which it is propagated in the dendrite¹².

Turrigiano *et al.*³ have observed up- and downregulation of the total postsynaptic strength, which does not operate in a synapse-specific manner. They compared AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) miniature excitatory postsynaptic current (mEPSC) amplitudes in 4–5-day cultures of visual cortex, bathed for tens of hours either with tetrodotoxin (to block spiking activity) or with a GABA_A antagonist, bicuculline (to increase firing rates). They found that periods of artificially increased activity lead to a decrease in synaptic gain, as predicted from competitive learning. In contrast, after artificially maintained periods of decreased activity, they observed the reverse effect — that is, an increase in synaptic gain.

Although the amplitude of mEPSCs was altered, neither their frequency nor their kinetics varied as a result of the past activity of the network. Thus, the observed effect is

due to changes in postsynaptic AMPA-receptor sensitivity, affecting spike-induced EPSCs as well as miniature synaptic events. This process, described in cultures of pyramidal neurons of rat visual cortex, differs from the supersensitivity that has been found at the neuromuscular junction. Here, the enhancement is extrajunctional, and no change in the amplitude of the miniature synaptic events is observed.

But interpretation of the results may be more complex than it seems. First, these observations have been made in a cultured system only. Certain forms of plasticity that are found only in cultures may signal to the rest of the network a failure of a local synaptic homeostatic regulation. Depression signals spread to other synapses from the site at which the depression is induced, possibly because the general state of the cultured network is too far from equilibrium¹³. Second, these results need to be validated relative to other processes that are also known to occur. For example, the development of multiple synaptic contacts is characteristic, and an inverse relationship between the number of synapses and quantal amplitude is found, independent of time¹⁴.

Another consideration is that after treatment with tetrodotoxin or bicuculline, morphological and electrophysiological changes have been found which depend on the activation of NMDA (*N*-methyl-D-aspartate) receptors. Although Turrigiano *et al.* used the NMDA-receptor antagonist AP5 (*D*(-)-amino-7-phosphonovaleric acid), this does not exclude the involvement of NMDA in inducing the effects. A more probing test would have been to associate AP5 treatment with the blockade of AMPA receptors rather than applying it alone, which here does not change the network's activity. Finally, the effect might be specific to the culture or cell type. In other cultures, opposite effects that support a presynaptic mechanism have been reported¹⁵.

These remarks should not diminish the importance of the new findings³. The form of synaptic plasticity described by Turrigiano *et al.* operates very differently from long-term potentiation and long-term depression. Moreover, because it is bidirectional, it may have interesting computational properties. Indeed, the scaling process may have two facets. The homeostasis side comes from the fact that multiplicative scaling of synaptic strengths will preserve relative differences between inputs, while maintaining the output of a given cell roughly constant. But at the same time, the activity requirements for inducing synaptic changes in the network will also be modified. When inputs are scaled up or down, co-activation of a smaller or larger set of afferents will be required to depolarize the postsynaptic cell sufficiently, or to trigger an increase in calcium levels to induce plasticity.

This view is reminiscent of the hypothesis

of a floating threshold in postsynaptic plasticity, proposed for the visual cortex on purely theoretical grounds¹⁶ and later worked out in more physiological terms¹⁷. The hypothesis holds that the postsynaptic plasticity threshold controlling the transition from long-term depression to long-term potentiation depends on the mean past activity of the cell. The plasticity threshold becomes easier or more difficult to reach by the cell's internal state, depending on whether the mean past neuronal activity — reflecting that of the whole network — has been reduced or increased, respectively. The mechanism described by Turrigiano *et al.* could play a similar role by scaling synaptic gain as a function of past activity. Linking the homeostasis of postsynaptic strength to functional plasticity remains the next challenge — and this will mean leaving the culture model for the more uncertain living brain. □

Yves Frégnac is in the Equipe Cognisciences, Institut Alfred Fessard, CNRS, 91198 Gif sur Yvette, France.
e-mail: Yves.Fregnac@iaf.cnrs.gif.fr

1. Hebb, D. O. *The Organization of Behavior: a Neurophysiological Theory* (Wiley, New York, 1949).
2. Frégnac, Y. in *Handbook of Brain Theory and Neural Networks* (ed. Arbib, M.) 459–464 (MIT Press, 1995).
3. Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. *Nature* **391**, 892–896 (1998).
4. Markram, H. & Tsodyks, M. *Nature* **382**, 807–810 (1996).
5. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. *Science* **275**, 220–224 (1997).
6. O'Donovan, M. J. & Rinzel, J. *Trends Neurosci.* **20**, 431–433 (1997).
7. Abraham, W. C. & Bear, M. F. *Trends Neurosci.* **19**, 126–130 (1996).
8. Glanzman, D. L., Kandel, E. R. & Schacher, S. *Neuron* **7**, 903–913 (1991).
9. Willshaw, D. G. & Von der Malsburg, C. *Proc. R. Soc. Lond. B.* **194**, 431–445 (1976).
10. Miller, K. D. *Neuron* **17**, 371–374 (1996).
11. Turrigiano, G., Abbott, L. F. & Marder, E. *Science* **264**, 974–977 (1994).
12. Chen, W. R., Midtgard, J. & Shepherd, G. M. *Science* **278**, 463–467 (1997).
13. Fitzsimonds, R. M., Song, H. J. & Poo, M. M. *Nature* **388**, 439–448 (1997).
14. Liu, G. & Tsien, R. W. *Nature* **375**, 404–408 (1995).
15. Tong, G., Malenka, R. C. & Nicoll, R. A. *Neuron* **16**, 1147–1157 (1996).
16. Bienenstock, E., Cooper, L. N. & Munro, P. J. *Neurosci.* **2**, 32–48 (1982).
17. Bear, M. F. *Neuron* **15**, 1–4 (1995).

Diabetes

A signal for β -cell failure

Joseph Avruch

Human type 2 diabetes is almost always accompanied by defects in both the responsiveness to insulin (which is commonly called insulin resistance) and the secretion of insulin. The result is hyperglycaemia — an excess of glucose in the bloodstream. Severe insulin resistance alone does not usually cause sustained hyperglycaemia, because the pancreatic β -cells can secrete much more insulin than is normally required. Moreover, in the face of insulin resistance, the β -cells have a substantial compensatory reserve. Although insulin resistance increases with the development of obesity and physical deconditioning, hyperglycaemia results mainly when the compensatory overproduction of insulin fails¹.

On page 900 of this issue, Withers *et al.*² describe the phenotype of mice that lack both of the alleles encoding the insulin-receptor substrate protein-2, IRS-2. These mice develop a syndrome that mimics the common form of human type 2 diabetes. As expected, the IRS-2 knockout mice are considerably insulin resistant. But, in contrast to other murine models with severe insulin resistance, which show a marked increase in β -cell mass, the IRS-2 knockout mice have a smaller mass of β -cells than wild-type mice. Thus, Withers *et al.* have not only verified the importance of IRS-2 in insulin action, but they have identified IRS-2 as one of the molecular determinants of β -cell compensation.

The IRS proteins were first detected as polypeptides (relative molecular mass, M_r , 160,000–185,000) that underwent rapid

tyrosine phosphorylation in response to insulin and insulin-like growth factor-1 (IGF-1), but not to other growth factors that act through receptor tyrosine kinases³. The IRS-1 protein contains an amino-terminal pleckstrin homology domain, followed by a phosphotyrosine-binding domain. Together, these enable binding to the activated insulin/IGF-1 receptor. The IRS proteins also have long carboxy-terminal tails that contain 8–18 potential tyrosine phosphorylation sites, many of which are in a YXXM motif — a preferred sequence for the insulin/IGF-1 receptor kinases. Once tyrosine phosphorylated (Fig. 1), IRS-1 binds proteins that contain SH2 domains, and, among this array, the phosphatidylinositol-3-OH kinase is probably the most important to the insulin regulation of energy metabolism.

Four IRS isoforms are currently known. The IRS-2 protein (M_r , 180,000) was first found in myeloid progenitor cells. IRS-3 (M_r , 60,000) is an adipocyte-specific IRS isoform⁴, and IRS-4 (M_r , 160,000) is a distinct homologue⁵. Because tyrosine phosphorylation of IRS-2 is promoted by interleukin-4 (as well as by insulin and IGF-1), the role of IRS proteins in the actions of other ligands whose receptors signal through Janus-family non-receptor tyrosine kinases was examined. The interferons, and many members of the interleukin-2- and interleukin-6-receptor families (including, for example, growth hormone), are now known to stimulate tyrosine phosphorylation of the IRS proteins³.

What functions the IRS proteins provide, that are not already supplied by the docking sites found on the receptor and non-receptor tyrosine kinases themselves, is not yet known.

So what happens when the IRS proteins are knocked out? IRS-1 knockout mice show a surprisingly modest phenotype^{6,7}. They have normal development at birth, but they are only 50–60% the weight of wild-type mice and remain comparatively small through life. Despite being clearly insulin resistant, these mice rarely develop fasting hyperglycaemia. Although insulin signal transduction is greatly reduced in skeletal muscle, it is maintained in the liver, and β -cell compensation is robust.

With such a relatively weak phenotype in homozygous knockout mice, one is inclined to discount the importance of heterozygous point mutations in the human IRS-1 gene. But functionally significant polymorphisms have been identified in the IRS-1 gene, and these are more frequent in patients with type 2 diabetes. Moreover, heterozygous deficiency of the IRS-1 and insulin-receptor genes in mice — although each alone is a silent event — together produce a strong diabetic phenotype⁸. The genetic basis for insulin resistance may, therefore, derive from the multiplicative effects of common polymorphisms at several steps in the insulin signalling pathway, each causing only a modest functional impairment.

In IRS-1 knockout mice, expression of IRS-2 is upregulated in skeletal muscle, suggesting that IRS-2 may compensate, in part, for the IRS-1 deficiency. Withers *et al.*² now establish IRS-2 not only as an important player in insulin signalling, but also as a candidate diabetes gene in man⁹. The unexpected element in the IRS-2 knockout phenotype is the failure of robust β -cell compensation. The compensatory increase in β -cell mass that occurs in insulin-resistant states reflects some combination of islet-cell hypertrophy, replication, neogenesis by differentiation from ductal precursors, and diminished apoptosis¹⁰.

How does a deficiency of IRS-2 curtail β -cell compensation? Withers *et al.* show that IRS-2 colocalizes with insulin in pancreatic islets, and that IRS-2 is expressed in ductal precursor cells. The β -cells also express IRS-1, insulin receptors and IGF-1 receptors. The signalling elements that are controlled by IRS-2 (such as phosphatidylinositol-3-OH kinase and Ras) are crucial regulators of mitogenesis, cellular differentiation and apoptosis. Identification of the downstream targets of these signals in the β -cell and its progenitors will be crucial to the understanding and manipulation of β -cell fate.

Pancreatic β -cells continue to replicate throughout a rodent's life, balanced by cell death, and gene knockouts of several transcription factors have provided an outline of

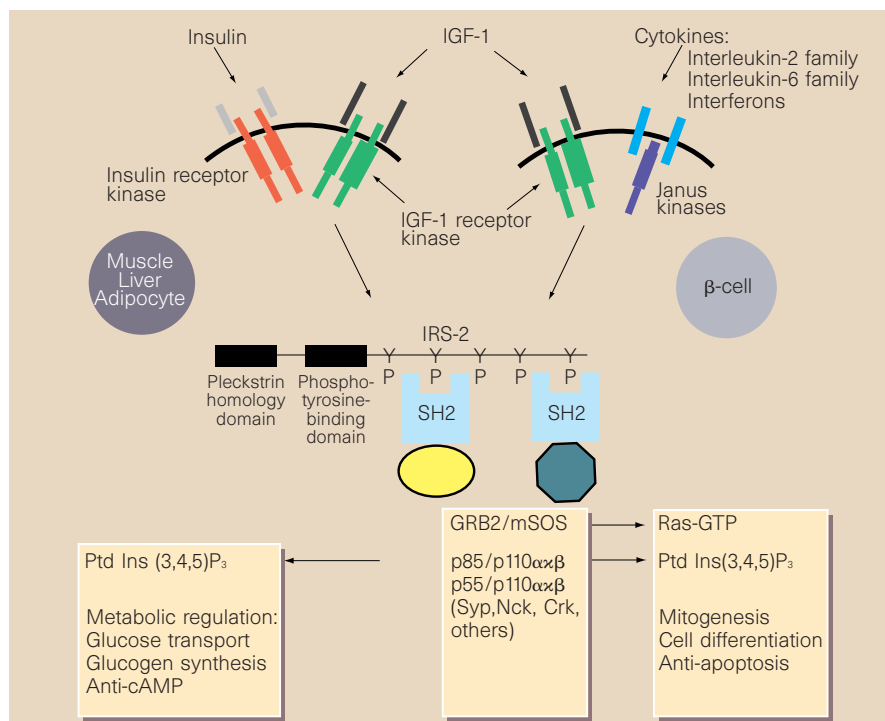


Figure 1 Signal transduction through IRS-2 in the β -cell and in insulin-sensitive tissues. Withers *et al.*² have found that *IRS-2* knockout mice develop a syndrome that closely resembles human type 2 diabetes, implicating *IRS-2* as a candidate diabetes gene. Insulin, IGF-1 and cytokines bind to their respective receptors, resulting in tyrosine phosphorylation of IRS-2 and recruitment of signalling proteins. These include the phosphatidylinositol-3-OH kinases and the Ras-specific guanyl nucleotide exchanger, mSOS, which are likely to be important for the role of IRS-2 in β -cell compensation to insulin resistance. Generation of phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃), and perhaps other signals, mediates the characteristic metabolic response to insulin in skeletal muscle, liver and adipose tissue.

the transcriptional basis for β -cell development¹¹. With the exception of hyperglycaemia, the signals that control the fate of β -cells — either during development or in adult life — are poorly defined. A 96-hour glucose infusion in the rat results in a 50% increase in β -cell mass, through increased replication and cellular hypertrophy¹². Nevertheless, the mediators and mechanisms of this response to glucose are not known. Better insight into the control of β -cell fate is necessary to understand what ordinarily limits β -cell compensation in the face of insulin resistance, and to what extent these limitations are genetically programmed or environmentally imposed. The success of β -cell replacement therapies will also depend on such knowledge. For example, long-term β -cell function is much greater when islets are transplanted into human recipients in the context of an intact pancreas, compared with isolated islets. This undoubtedly relates to the capacity for β -cell neogenesis, and the ability of the β -cells to replicate and/or to avoid apoptosis under oxidative or anoxic stress. The contribution of IRS-2 and its partners to these processes will now come under scrutiny. □

Joseph Avruch is in the Diabetes Unit, Department of Medicine, and the Department of Molecular Biology, Massachusetts General Hospital and

Harvard Medical School, 50 Blossom Street, Boston, Massachusetts 02114, USA.

e-mail: avruch@helix.mgh.harvard.edu

- Lillioja, S. *et al.* *N. Engl. J. Med.* **329**, 1988–1992 (1993).
- Withers, D. J. *et al.* *Nature* **391**, 900–904 (1998).
- Yenush, L. & White, M. F. *BioEssays* **19**, 491–500 (1997).
- Lavan, B. E., Lane, W. S. & Lienhard, G. E. *J. Biol. Chem.* **272**, 11439–11443 (1997).
- Lavan, B. E. *et al.* *J. Biol. Chem.* **272**, 21403–21407 (1997).
- Tamemoto, H. *et al.* *Nature* **372**, 182–186 (1994).
- Araki, E. *et al.* *Nature* **372**, 186–190 (1994).
- Bruning, J. C. *et al.* *Cell* **88**, 561–572 (1997).
- Elbein, S. C. *J. Nutr.* **127**, 1891S–1896S (1997).
- Finegood, D. T., Scaglia, L. & Bonner-Weir, S. *Diabetes* **44**, 249–256 (1995).
- Habener, J. F. & Stoffers, D. A. *Proc. Assoc. Am. Phys.* **110**, 12–21 (1998).
- Bonner-Weir, S., Deery, D., Leahy, J. L. & Weir, G. C. *Diabetes* **37**, 232–236 (1989).

Errata

In the News and Views Feature “Galileo at Jupiter — meetings with remarkable moons” by William B. McKinnon (*Nature* **390**, 23–26; 1997), the heat flow from Io was given as 2.5 mW m⁻². It is actually 2.5 W m⁻², to compare with a terrestrial average of 90 mW m⁻² (Pollack, H. N., Hurter, S. J. & Johnson, J. R. *Rev. Geophys.* **31**, 267–280; 1993). ‘Size’ in Table 1 refers to radius.

In the News and Views article “Fixed hotspots gone with the wind” by Ulrich Christensen (*Nature* **391**, 739–740; 1998), the Web-site address in reference 9 was misspelt. It should have read: <http://www.ngdc.noaa.gov/mgg/image/seafloor.html>

Daedalus

Cold porridge

Last week Daedalus presented his new Hibernator. The subject is cooled down strongly enough to stop his heart, but his blood is kept moving by rhythmic hydrostatic pressures applied in sequence to his limbs and regions of his body. Thus his vital organs still receive their greatly reduced demand for oxygen and fuel. Monitoring equipment keeps him stable; at wake-up time, microwave heating swiftly restores his normal body temperature.

The primary beneficiaries of the technology are old people. A municipal Hibernatorium could see thousands of pensioners safely through the winter at far less cost than normal pension and social service provisions. Not only would they dodge the hardships of cold weather; they would live longer. Ageing, and the stealthy advance of degenerative diseases, will almost cease at hibernation temperatures. A pensioner with ten years of life in him could last for twenty years — and all that time would be summer. He would not need to relocate to some warm but dreary southern geriatric resort; he could stay around to follow the progress of his grandchildren, and experience and deplore what the world is coming to, for a much longer period.

The mental effects of the Hibernator remain to be explored. Will six months of hibernation seem like six months to the memory? If not, hibernating pensioners will hardly notice their periods of torpor. But if so, the Hibernator could revolutionize the prison service as well. Most prisoners would be happy to serve their time unconscious; and most prison governors would be only too pleased to let them. At the end of (say) ten years’ hibernation, the criminal would wake up to recall his crime and late way of life as a distant folly. He could reasonably vow to leave all that behind, and start life afresh. Unconsciousness would be a far better form of rehabilitation than sharing a gaol with an ever-changing crew of lively expert criminals, all keen to swap ideas.

The Hibernator could serve still other purposes. The long-term unemployed could sleep out an economic downturn; unrequited lovers might embrace it as a way to forget. Those valiantly predicting a religious revival, or the resurgence of Socialism, or invasion by aliens, could use it to await (or sleep safely through) these social revolutions. And the younger child’s hopeless threat to his big brother — “Wait till I’m older than you!” — would at last have a chance of coming true.

David Jones

Obituary

Masaru Ibuka (1908–97)

Electrical engineer and co-founder of SONY

On 16 December 1947 a group of scientists at Bell Laboratories in the United States, including the eventual Nobel laureates John Bardeen, William Shockley and Walter Brattain, announced that they had developed the transistor. But the device was neither ready for production on industrial scales, nor could it operate at high enough frequencies for most applications. These and other technical problems had to be surmounted before the transistor could fulfil its immense potential. Equally importantly, visionary engineers and entrepreneurs had to recognize the significance of the invention before its time had really come, and attract, organize and inspire the right people to build popular and profitable products, initially using very limited resources.

In Japan, two such engineers and entrepreneurs were the co-founders of SONY, Akio Morita and Masaru Ibuka. Ibuka died on 19 December 1997, almost 50 years to the day after that announcement from Bell Labs and the debut of the invention which he was to help develop so far.

Ibuka was born on 11 April 1908, in the beautiful town of Nikko in Tochigi Prefecture, about 100 km north of Tokyo. His was a technologically aware family — his father had studied electrochemistry at the school which has now become the famous Tokyo Institute of Technology, and as a young man he constructed one of Japan's first water-driven electrical power stations.

In 1910, following his father's tragic accidental death by electrocution, Ibuka moved with his mother to live with his paternal grandparents in Kobe. After his mother's remarriage, Ibuka stayed with his grandparents and was indulged in his fascination with technology and science. For example, it is said that he frequently took alarm clocks to pieces, and when he could not fit them together again his grandfather would buy him a new one to start again. Pursuing that interest into higher education, in 1933 Ibuka graduated in electrical engineering from Waseda University.

After graduation, Ibuka joined the Photochemical Laboratory, where he worked on sound and picture recording technologies. In 1936 he moved to the Japan Opto-acoustic Engineering Company, where he worked on vacuum tubes. In 1940 the vacuum tube department



became independent and Ibuka joined the Japan Measurement Instrument Company, where he developed military technology, such as submarine detection instruments.

When the war ended, Ibuka decided to concentrate on consumer products. In 1945, in the devastated centre of Tokyo, he founded the Tokyo Tsushin Kenkyusho (the Tokyo Communications Laboratory). In 1946, Ibuka, Morita and Maeda converted this laboratory into the Tokyo Tsushin Kogyo K. K. (the Tokyo Telecommunications Engineering Corporation, which in 1958 became SONY). With sometimes stubborn insistence Ibuka would hire the best people he could find: directors of government laboratories and engineers of the Japanese public radio corporation NHK, and even those of other companies. He is reported to have phoned his chosen candidate, and if necessary the candidate's superiors and colleagues, until he got his man.

Ibuka had a clear vision of new products, and made up his mind over resource investment very quickly. Decisions to transfer potential products out of the laboratory and into production development were taken on the spot and implemented within one or two weeks, including the necessary personnel transfers. This combination of vision and decisiveness had an immediate and stunning effect on his engineers, whose enthusiasm for the company was such that they drove themselves hard in the quest for success.

A well-formed idea of the products to be developed and of the quality required, and top management involvement in the process, were key elements in the success of Ibuka and Morita's company. Thus Ibuka was the leader, or heavily involved in the development teams, for many remarkable innovations. Among them were the magnetic recording tape in 1949; a tape recorder in 1950, and later the video tape

recorder; Japan's first transistor radio in 1955; a pocket-size transistor radio in 1957; the world's first transistor television in 1960; and the Trinitron colour television system in 1967.

Ibuka was a central figure in the reconstruction of Japan after the disastrous consequences of the Second World War. Part of the success of the Tokyo Tsushin Kogyo K.K. may well have stemmed from General Douglas MacArthur's initiatives to re-establish Japan's broadcasting and telecommunications structure as quickly as possible. Certainly, Ibuka is said to have been a keen student of two young American engineers, Homer Sarasohn and Charles Protzman, who as part of the reconstruction programme taught modern quality management methods in Japan.

For Japan and the Japanese people, SONY is an unusual company and Ibuka was an unusual man. Ibuka was a Christian (a Protestant) in a country where the overwhelming majority of the population regularly attends both Buddhist and Shintoist religious ceremonies, shrines and temples. SONY has probably always been the least Japanese of all the principal Japanese companies in its corporate governance and the internationalization of its operations, and even in its scientific creativity — Leo Esaki was awarded one of Japan's rare Nobel prizes for his discovery of tunnelling phenomena in semiconductors, which he made while employed at Ibuka's Tokyo Tsushin Kogyo K.K.

Like Bill Hewlett and David Packard in the United States, Morita and Ibuka built a remarkable corporation. New-style organizations such as these explore fresh ways for creative people to work together, and company cultures which provide the fertile soil for further innovation and progress. These companies survive the founders and continue to grow and adapt: even during the present Asian financial crisis, Ibuka's SONY seems to thrive and grow, unlike several of its competitors.

Ibuka's talents extended to writing, and he published several books, including one about his friend Souichiro Honda — the founder of the Honda Motor Company, and another remarkable Japanese entrepreneur to emerge after the Second World War. To my knowledge these books are only available in Japanese, so that Masaru Ibuka is not as well known outside Japan as he deserves. But SONY, the company that he co-created, surely is.

Gerhard Fasol

Gerhard Fasol is at Eurotechnology Japan K.K., Parkwest Building 11th Floor, 6-12-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 160-0023, Japan. e-mail: fasol@euro-technology.com

Latham's life-forms

William Latham is working to establish a place for computers in the world of art. He sets the design rules to be followed and his program generates images in a process that could be called a simulation of evolution in action.

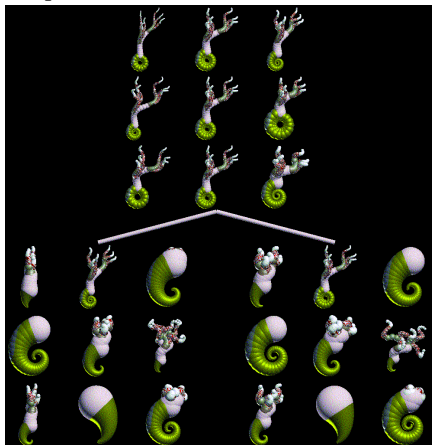
Martin Kemp

It seems to me that there is a real problem with computer art. The problem does not reside in getting the new tool to perform old artistic tricks but rather with a more profound identification of what new territories computers and art might inhabit in terms of the inherent nature of computational procedures.

One approach, pioneered by Benoit Mandelbrot in his fractal images, has been to use iteration to model mathematical sets in visual terms. The results may be seen as 'art' — just as aesthetic status can be accorded to Renaissance drawings of the Platonic solids or to Henri Poincaré's geometrical models — but they inhabit only a limited domain in the wide territories already colonized by the visual arts.

Standard computer-aided design programs, and the kinds of software 'palettes' aimed specifically at artists, are at best seen as providing ingenious extensions to hand-driven techniques — or at worst as giving lazy operatives the opportunity to exploit methods they don't understand. But a few artists are conducting a more searching exploration of art forms that can be generated only with computers.

The most promising approach, exemplified by British sculptor William Latham, is to set up programs on the basis of aesthetic choices in such a way that the parameters of style and content in the images are established but the final form is not predetermined. Indeed, a fixed 'final form' may not be the end in view. Rather, the evolutionary program gives rise to a new type of kinetic sculpture.



Latham's *Family Tree of Evolving Forms*, computer-generated using Mutator.



Latham's *Evolving Form*, computer-generated using Mutator.

Latham spent six years as research fellow in the IBM Scientific Centre at Winchester in southern England, and has developed his programs in collaboration with Stephen Todd. Most notable is a powerful and wide-ranging design tool called Mutator, with which sculptural forms are endowed with genetic properties that shape their growth. The basis of Mutator is provided by Form Grow, a geometrical grammar that uses spirals and fractal recursion to emulate the kinds of geometry of natural forms that have fascinated students of nature from the time of Leonardo and Dürer.

Random mutation allows the artist to navigate through the space of the infinitely varied forms that are inherent in Form Grow. The program also provides rules through which the 'life-forms' are subject to processes of 'natural selection'. The results of such "Darwinian evolution driven by human aesthetics" are fantastical organisms whose morphologies metamorphose in a sequence of animated images.

The still images on this page show the kinds of form Latham generates, working "like a 'gardener' or 'farmer', repeatedly picking, marrying and breeding, starting from a simple structure to evolve thousands of complex genetic variations". They are intricately wondrous in their compulsive convolutions, combining a strange beauty with an air of predatory menace.

Latham's images are typical of those generated by computer in that they seem irredeemably to have a computerized 'look' about them. There is a characteristic visual feel — just as the look of an oil painting is reliant on the properties of that particular medium.

Latham openly exploits these qualities, not least in relation to a certain kind of 'science-fiction' aesthetic that owes more than a little to the imaginative visions of the best draughtsmen of children's comics and the designers of Hollywood films. It is one of the characteristic visual modes of our age. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Skin immunization made possible by cholera toxin

Immunization using an application to the skin surface, without physical penetration by needles, would greatly increase the ease of vaccination. In orally and nasally administered vaccines, the bacterial product cholera toxin (CT) is commonly used to enhance the immune response. We found that when CT was applied to the surface of the skin, it stimulated an immune response to vaccine components such as diphtheria or tetanus toxoids. Immunization can thus be achieved by the simple application of a mixture of CT and vaccine components without penetration or disruption of the skin.

CT is responsible for the symptoms of cholera¹, is produced by the bacterium *Vibrio cholerae* and is a member of the adenosine diphosphate (ADP)-ribosylating bacterial exotoxins². CT is already widely used experimentally as an adjuvant³, a compound used to enhance the immune response to vaccine components. Given by the oral or nasal route with proteins⁴ or vaccine components such as tetanus toxoid⁵,

CT and its derivatives enhance the resulting antibody responses.

Transcutaneous immunization (immunization using the skin) is achieved by simple application of a solution to the surface of the skin. To gain access to mouse skin, we shaved the animals' backs with a clipper and allowed them to rest for 24 hours before immunization. To avoid the possibility of oral immunization occurring through grooming, we anaesthetized the mice and then wetted the skin surface with immunizing solution for 120 minutes. Finally, we washed the skin extensively with water. No redness or swelling was observed at the immunization site for up to 72 h after the primary or boosting immunizations, and skin biopsies at the site of exposure showed no signs of inflammation. In separate experiments to exclude the possibility of immunization secondary to skin disruption, we also immunized mice on an unshaved ear.

Transcutaneous immunization was first

achieved using CT by itself. When given alone by the oral route, CT stimulates a potent immune response in the form of anti-CT antibodies⁶. We tested the capacity of CT to induce a similar response through the skin by applying a saline solution of CT to the bare skin of the shaved mouse. This application induced high levels of IgG antibodies specific for CT (Fig. 1a).

The high levels of antibodies produced against CT administered by the transcutaneous route suggested that CT might also act as a skin adjuvant to enhance the immune response to proteins and vaccine components placed on the skin.

Bovine serum albumin (BSA) is a large protein antigen of the type that, when given by the mucosal route, requires an adjuvant such as CT to induce an immune response³. We achieved induction of BSA-specific antibodies only when BSA was coadministered with CT (Fig. 1b). Thus, CT acted as a transcutaneous adjuvant, promoting the immune response to BSA. CT also acted as a transcutaneous adjuvant for the common vaccine components diphtheria toxoid and tetanus toxoid (Fig. 1c,d).

A critical role in enhancing the immune response to both licensed and experimental vaccines is played by adjuvants⁷. These compounds, and especially CT and related products, have been important in the development of the mucosal route as a useful and easily accessible non-invasive way to induce immunity³. Employing CT as an adjuvant for transcutaneous immunization would allow the skin to be used in a similar way. This method could be particularly useful given the large surface area of the skin, its accessibility and the existence of potent immune cells within it⁸.

Skin immune responses can also be induced by skin disruption with depilatory agents⁹ or by scarification as in smallpox inoculation¹⁰, but transcutaneous immunization involving CT offers the benefit of leaving the skin intact. Using CT to induce immune responses through the skin to common off-the-shelf vaccine components — such as tetanus toxoid and diphtheria toxoid — may ultimately lead to the development of simple and safe needle-free vaccines.

**Gregory M. Glenn, Mangala Rao,
Gary R. Matyas, Carl R. Alving**
Department of Membrane Biochemistry,
Walter Reed Army Institute of Research,
Washington, DC 20307-5100, USA

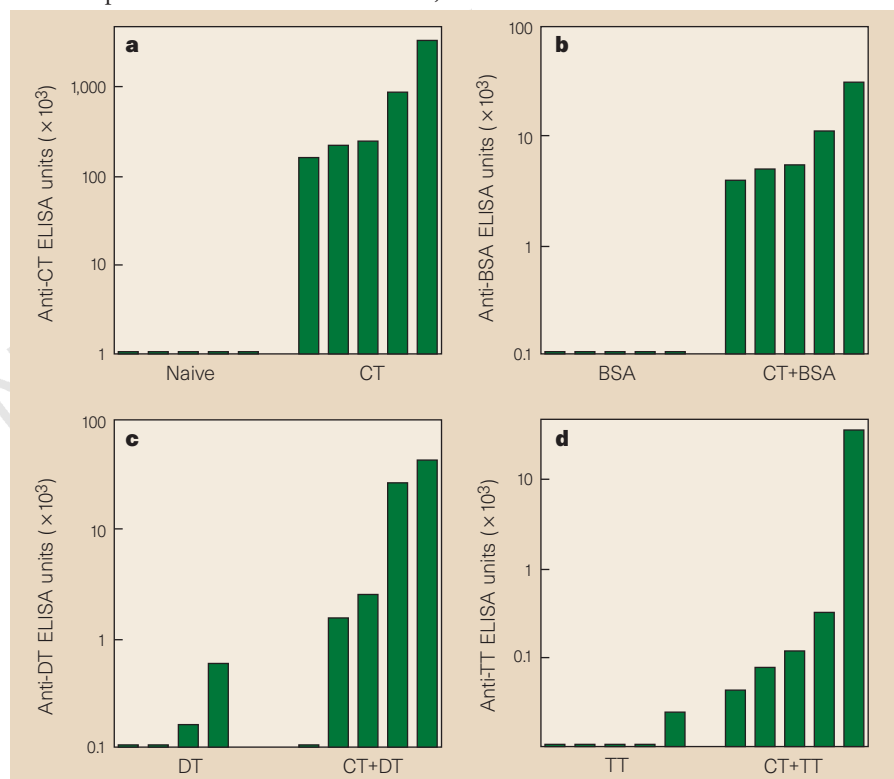


Figure 1 Individual IgG antibody responses to antigens using cholera toxin (CT) as adjuvant. BALB/c mice ($n=5$) were immunized transcutaneously with a saline solution of: **a**, 100 μ g CT; **b**, 200 μ g of bovine serum albumin (BSA), or 200 μ g of BSA and 100 μ g of CT; **c**, 100 μ g of diphtheria toxoid (DT) or 100 μ g of DT and 100 μ g of CT; **d**, 50 μ g of tetanus toxoid (TT) or 50 μ g of TT and 100 μ g of CT. Mice were immunized and boosted at 8 weeks (a, c, d) or 3 weeks (b) later. Sera were analysed for IgG antibodies by solid-phase ELISA two weeks after the second immunization and compared to pre-immune sera. Antigen-specific IgG responses were determined and are reported as the geometric mean of individual sera in ELISA units, the inverse dilution at which the absorbance (405 nm) is equal to 1.0.

- De, S. N. *Nature* 183, 1533–1534 (1959).
- Krueger, K. M., Barbieri, J. T. *Clin. Microbiol. Rev.* 8, 34–47 (1995).
- Snider, D. P. *Crit. Rev. Immunol.* 15, 317–348 (1995).
- Lycke, N., Holmgren, J. *Immunology* 59, 301–308 (1986).
- Jackson, R. J. *et al. Infect. Immun.* 61, 4272–4279 (1993).
- Elson, C. O., Ealding, W. J. *Immunol.* 132, 2736–2741 (1984).
- Edelman, R. *Rev. Infect. Dis.* 2, 370–383 (1980).
- Bos, J. D. *Clin. Exp. Immunol.* 107, (Suppl.1), 3–5 (1997).
- Tang, D., Shi, Z., Curiel, D. T. *Nature* 388, 729–730 (1997).
- Benenson, A. S. *JAMA* 143, 1238–1240 (1950).

Rescuing *Wolbachia* have been overlooked...

Wolbachia, intracellular bacteria transmitted through the egg, have been estimated to infect more than 16% of all insect species, as well as other arthropods¹. They distort their hosts' reproduction, inducing parthenogenesis, feminization and cytoplasmic incompatibility². This favours the reproduction of infected female hosts at the expense of uninfected females. Here we show that several *Wolbachia* strains that cannot generate modifications in host sperm can still rescue the modifications caused by other strains as long as the two strains are sufficiently closely related.

Cytoplasmic incompatibility causes embryonic mortality in crosses between infected male hosts and females that are either uninfected or infected with an unrelated *Wolbachia* strain. Cytological studies suggest that *Wolbachia* modifies the paternal chromosomes during spermatogenesis and this modification is rescued in eggs of females infected with the same strain of *Wolbachia* during fertilization^{3–5}.

In *Drosophila* two types of strains of *Wolbachia* have been described: $\text{mod}^+\text{resc}^+$, capable of modifying spermatozoa and rescuing this modification in eggs^{5,6}, and $\text{mod}^-\text{resc}^-$, incapable of either modifying sperm or rescuing the modification in eggs^{7–9}. It has been hard to explain the presence of $\text{mod}^-\text{resc}^-$ strains in insect hosts as it is difficult to understand how these strains could spread into host popula-

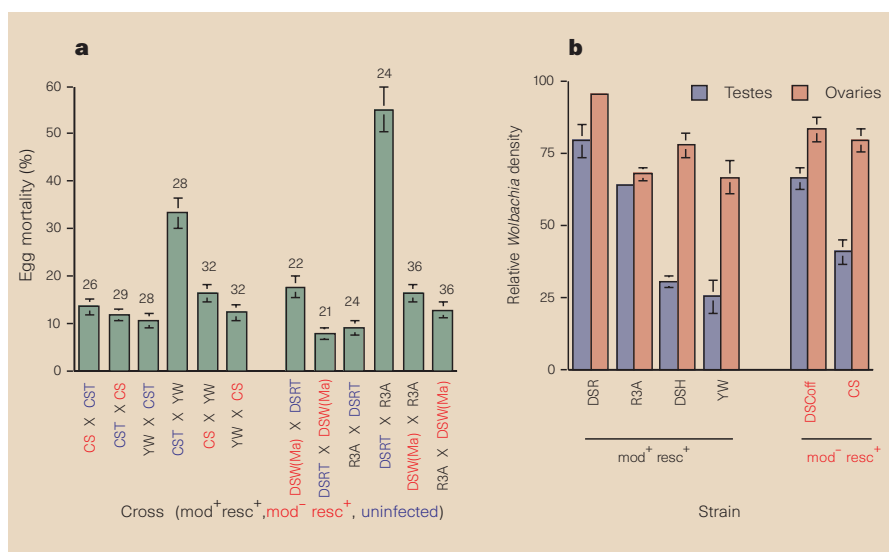


Figure 2 Phenotypes of *Wolbachia* crosses. **a**, Mean egg mortality in single-pair crosses (female × male) of mod^+ , mod^- and tetracycline-cured strains of *D. simulans* and *D. melanogaster* (values above bars represent number of single-pair replicates; error bars show s.e.m.). **b**, Mean relative *Wolbachia* densities in ovary and testis tissues of different *Drosophila* strains as determined by dot blot hybridization with a *Wolbachia* *dnaA* probe¹⁰⁶ (five replicates per strain; error bars show s.e.m.). Strain abbreviations are as in Fig. 1.

tions without the action of cytoplasmic incompatibility.

Alternatively these $\text{mod}^-\text{resc}^-$ strains might have failed to rescue cytoplasmic incompatibility in test crosses because of strain differences between the mod^+ and mod^- strains, as in cases of bidirectional incompatibility^{5,10}. Until recently there was no way to type *Wolbachia* strains independently of crossing the hosts carrying the strains. We have recently cloned and sequenced the gene (*wosp*) for a major surface protein of *Wolbachia* that shows variability between *Wolbachia* strains¹¹. The sequence of this gene from different *Wolbachia* strains infecting *Drosophila* shows that many of the crosses previously used to test the ability of mod^- strains to rescue cytoplasmic incompatibility were between distantly related *Wolbachia* variants.

We have crossed females from previously described $\text{mod}^-\text{resc}^-$ strains with males that, on the basis of *wosp* gene sequences, carry closely related $\text{mod}^+\text{resc}^+$ *Wolbachia* strains (Fig. 1) and find that the $\text{mod}^-\text{resc}^-$ strains can indeed rescue the sperm modification of these related *Wolbachia* strains. For example, the *Wolbachia* strains infecting *D. melanogaster* Canton-S and yw^{67C23} are closely related by *wosp* sequences, and crosses between Canton-S females and yw^{67C23} males are completely compatible (Fig. 2a). Similarly, the mod^- strain of *Wolbachia* originally described in *D. mauritiana* and subsequently transferred to the uninfected *D. simulans* Watsonville strain did not express cytoplasmic incompatibility in crosses with either uninfected females or females of the *D. simulans* Riverside strain infected with a *Wolbachia* variant known to display high levels of cytoplasmic incompatibility⁹.

Indeed, transinfected Watsonville females fail to rescue the cytoplasmic incompatibility phenotype in crosses with Riverside males⁹, but they were fully compatible in crosses with males of the R3A strain (Fig. 2a), which carries the $\text{mod}^+\text{resc}^+$ Noumea *Wolbachia* variant, a strain closely related by *wosp* gene sequences (Fig. 1). In similar crosses we found that the mod^- *Wolbachia* strain described in *D. simulans* populations from Coffs Harbour, Australia⁷, rescues the cytoplasmic incompatibility imprint of the *Wolbachia* $\text{mod}^+\text{resc}^+$ variant infecting *D. melanogaster* after transfer into *D. simulans* (unpublished data).

These strains, which fail to modify sperm but can rescue the modification in eggs, might do so because they show differential affinities for male and female germline tissues, being more abundant in eggs and less abundant in testes. We tested this hypothesis with dot blots of nucleic acids extracted from testes and ovaries, with the *Wolbachia* *dnaA* gene as a probe¹² (Fig. 2b) and by western blots of protein extracts from the same tissues with a polyclonal anti-WOSP antiserum (data not shown).

All $\text{mod}^+\text{resc}^+$ and $\text{mod}^-\text{resc}^+$ *Wolbachia* strains infected the testes and ovaries of their respective hosts, infecting ovaries at higher levels than testes, but there was no simple correlation between the absolute densities of *Wolbachia* in testes and the ability of these strains to act as either a mod^+ or a mod^- strain. With at least one of these strains the mod^- phenotype is independent of the host genome⁹, so mod^- *Wolbachia* variants might be genetically incapable of imprinting *Drosophila* sperm. Thus the molecular mechanisms of sperm

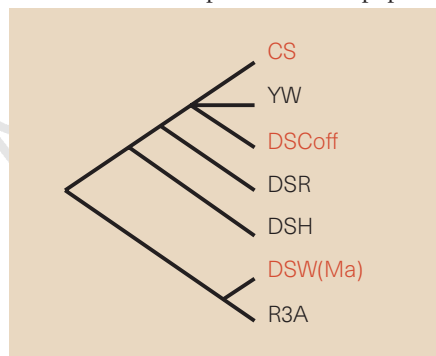


Figure 1 Single most parsimonious tree of *Wolbachia* strains infecting *Drosophila* hosts. The tree is based on a *Wolbachia* *wosp* gene data set¹¹ and shows the relationships between $\text{mod}^+\text{resc}^+$ strains (black) and $\text{mod}^-\text{resc}^+$ strains (red) (consistency index 0.922). YW, *D. melanogaster* yw^{67C23} (ref. 15); CS, *D. melanogaster* Canton-S⁸; CST, tetracycline-treated CS strain; DSR, *D. simulans* Riverside, CA⁶; DSRT, tetracycline-treated DSR strain; DSW(Ma), *D. simulans* Watsonville strain transinfected with the *D. mauritiana* *Wolbachia* symbiont⁶; R3A, *D. simulans* infected with the *Wolbachia* Noumea symbiont¹⁶; DSH, *D. simulans* Hawaii⁶; DSCoff, *D. simulans* Coffs Harbour strain⁷.

modification and rescue might involve multiple *Wolbachia* genes.

The existence of *Wolbachia* strains that do not modify sperm but can rescue the cytoplasmic incompatibility phenotype has been predicted by theory^{13,14}. Our work suggests that previously classified *mod⁻resc⁻* strains of *Wolbachia* can in fact rescue cytoplasmic incompatibility and can therefore be considered *mod⁻resc⁺*. This finding helps explain how *Wolbachia* *mod⁻* strains could have spread into insect populations without the action of cytoplasmic incompatibility. *Wolbachia* strains that can rescue but not cause cytoplasmic incompatibility can be expected to increase in frequency in populations by essentially parasitizing the cytoplasmic incompatibility sperm modification effect of a related *Wolbachia* strain.

The potentially lower fitness costs and perfect maternal transmission of these strains^{7,9} might explain how they could displace the initial strain that they parasitized in order to invade the host population. This is in accordance with models predicting that *Wolbachia* will evolve towards neutrality with respect to cytoplasmic incompatibility as long as maternal transmission frequency can be increased¹³.

Kostas Bourtzis*, Stephen L. Dobson, Henk R. Braig, Scott L. O'Neill

Section of Vector Biology,
Department of Epidemiology and Public Health,
Yale University School of Medicine,
60 College Street, New Haven, CT 06510, USA
e-mail: scott.oneill@yale.edu

*Alternative address: Institute of Molecular Biology
& Biotechnology, Foundation for Research and
Technology — Hellas, Heraklion, Crete, Greece

...and discovered on Mount Kilimanjaro

The endocyttoplasmic bacterium *Wolbachia* causes the death of arthropod embryos, when present in reproductive organs, by cytoplasmic incompatibility¹. *Wolbachia* is harboured in both sexes but transmission is maternal only. Cytoplasmic incompatibility occurs when infected males inseminate uninfected females² or females bearing a different variant of *Wolbachia*^{3,4}. Two kinds of *Wolbachia* have been described¹: (*mod⁺resc⁺*) which induces cytoplasmic incompatibility by modifying sperm but can rescue this phenotype when in the egg, and (*mod⁻resc⁻*) which neither induces cytoplasmic incompatibility nor rescues from it. Theory predicts a third kind (*mod⁻resc⁺*) that would not induce cytoplasmic incompatibility but would rescue from it⁵⁻⁷. We have found such a *Wolbachia* variant in a *Drosophila simulans* population on Mt Kilimanjaro, Tanzania. The existence of a *mod⁻resc⁺* *Wolbachia* shows that the modification and rescue functions can be dissociated with regard to the cytoplasmic incompatibility process.

We detected this *Wolbachia* variant (*wKi*) in isofemale lines from flies captured in March 1996. Of 49 lines, 9 showed a positive PCR signal with primers for the *Wolbachia* 16S rRNA gene⁸. Infected (*Kili⁺*) and uninfected (*Kili⁻*) Kilimanjaro males were crossed with *Kili⁻* females. No cytoplasmic incompatibility was detected. The percentage of unhatched eggs did not differ significantly whether the males were infected or uninfected ($12.3 \pm 2.6\%$ and $10.7 \pm 3.3\%$, respectively).

We tested the compatibility of *Kili⁺* flies with individuals infected by each of the three natural *mod⁺resc⁺* *Wolbachia* strains (*wRi*, *wHa* or *wNo*)⁹. *Kili⁺* males were compatible with females infected by *wRi* ($12.0 \pm 5.1\%$ unhatched eggs), *wHa* ($14.0 \pm 4.1\%$) or *wNo* ($8.5 \pm 1.3\%$). *Kili⁺* females were incompatible with males infected by *wRi* (100.0% unhatched eggs) or *wHa* ($84.7 \pm 3.3\%$). Surprisingly, the *Kili⁺* females were completely compatible with *wNo*-infected males ($8.5 \pm 2.1\%$ unhatched eggs), whereas *Kili⁻* females were incompatible ($70.2 \pm 3.5\%$). *wKi* thus behaves as a *mod⁻resc⁺* variant with regard to the *wNo* type.

Cytoplasmic incompatibility helps *Wolbachia* in invading previously uninfected populations, because it eliminates uninfected eggs, giving a higher fitness to infected females. However, during the invasion process, any *mod⁻* variant arising by mutation will also increase in frequency as long as it is still *resc⁺*. Under certain conditions,

these *mod⁻resc⁺* variants would completely replace the *mod⁺resc⁺* wild type. If this happened, the population would again become vulnerable to invasion by uninfected cytoplasm, because *mod⁻* mutants do not exert any selective pressure against uninfected eggs. The infection might then be lost with time. Meanwhile, because *mod⁺* variants have been eliminated from the population, the rescue capability becomes useless and *resc⁻* mutants can be expected to arise.

The *Wolbachia* variants discovered in wild *D. simulans* populations^{9,10} have been of two types: *mod⁺resc⁺* (*wRi*, *wHa*, *wNo*) and *mod⁻resc⁻* (*wMa*, *wA*). *wA* has been found both on the American and Australian continents, whereas *wMa* is known only in Madagascar. We established that *wA* and *wKi* are different because *wA* cannot rescue from the *wNo* cytoplasmic incompatibility phenotype¹⁰. The status of *wMa* is unknown in this respect, so *wMa* and *wKi* might be the same variant. Moreover, the 16S rDNA amplification product of *wKi* has one *VspI* restriction site, as do *wMa* and *wNo* sequences⁸, but such a site is absent from *wRi*, *wHa* and *wA* sequences.

Eastern Africa is thought to be the centre of origin of *D. simulans*¹¹, so the *Wolbachia* infection of the Kilimanjaro population could be one of the oldest in this species, implying that *mod⁻resc⁺* variants derive from a *mod⁺resc⁺* ancestor with time.

In contrast, *wNo* is known only in the Seychelles and in New Caledonia¹⁰, so perhaps *wNo*-infected populations originate from migrants that left East Africa bearing the *mod⁺resc⁺* ancestor. After geographic isolation, the Indo-Pacific variant could have remained *mod⁺resc⁺* while the ancestral variant has evolved into *mod⁻resc⁺* in Africa. The maintenance of the rescue function implies that it might have a further role.

The proportion of insect species potentially infected by *Wolbachia* is estimated at 16% (ref. 1), so further research might well yield other *mod⁻resc⁺* variants. Mt Kilimanjaro is probably just the tip of an iceberg.

Hervé Mercot, Denis Poinot

Laboratoire Dynamique du Génome et Évolution,
Institut Jacques Monod, CNRS-Université Paris 7,
75251 Paris, France
e-mail: mercot@ccr.jussieu.fr

- Werren, J. H., Windsor, D. & Guo, L. R. *Proc. R. Soc. Lond. B* **262**, 197–204 (1995).
- O'Neill, S. L., Hoffmann, A. A. & Werren, J. H. *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction* (Oxford Univ. Press, Oxford, 1997).
- Callaini, G., Dallai, R. & Riparbelli, M. G. *J. Cell Sci.* **110**, 271–280 (1997).
- Callaini, G., Riparbelli, M. G., Giordano, R. & Dallai, R. *J. Invert. Pathol.* **67**, 55–64 (1996).
- O'Neill, S. L. & Karr, T. L. *Nature* **348**, 178–180 (1990).
- Hoffmann, A. A., Turelli, M. & Simmons, G. M. *Evolution* **40**, 692–701 (1986).
- Hoffmann, A. A., Clancy, D. & Duncan, J. *Heredity* **76**, 1–8 (1996).
- Holden, P. R., Jones, P. & Brookfield, J. F. *Genet. Res.* **62**, 23–29 (1993).
- Giordano, R., O'Neill, S. L. & Robertson, H. M. *Genetics* **140**, 1307–1317 (1995).
- Montchamp-Moreau, C., Ferveur, J. & Jacques, M. *Genetics* **129**, 399–407 (1991).
- Zhou, W., Rousset, F. & O'Neill, S. L. *Proc. R. Soc. London B* (in the press).
- Bourtzis, K., Nirgianaki, A., Onyango, P. & Savakis, C. *Insect Mol. Biol.* **3**, 131–142 (1994).
- Turelli, M. *Evolution* **48**, 1500–1513 (1994).
- Werren, J. H. & O'Neill, S. L. in *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction* (eds O'Neill, S. L., Hoffmann, A. A. & Werren, J. H.) 1–41 (Oxford Univ. Press, Oxford, 1997).
- Bourtzis, K., Nirgianaki, A., Markakis, G. & Savakis, C. *Genetics* **144**, 1063–1073 (1996).
- Mercot, H., Llorente, B., Jacques, M., Atlan, A. & Montchamp-Moreau, C. *Genetics* **141**, 1015–1023 (1995).

- Werren, J. H. *Annu. Rev. Entomol.* **42**, 587–609 (1997).
- Hoffmann, A. A., Turelli, M. & Simmons, G. M. *Evolution* **40**, 692–701.
- Breeuwer, J. A. J. & Werren, J. H. *Nature* **346**, 558–560 (1990).
- O'Neill, S. L. & Karr, T. L. *Nature* **348**, 178–180 (1990).
- Prout, T. *Evolution* **48**, 909–911 (1994).
- Turelli, M. *Evolution* **48**, 1500–1513 (1994).
- Hurst, L. D. & McVean, G. T. *Proc. R. Soc. Lond. B* **263**, 97–104 (1996).
- Rousset, F., Vautrin, D. & Solignac, M. *Proc. R. Soc. Lond. B* **247**, 163–168 (1992).
- Clancy, D. J. & Hoffmann, A. A. *Trends Ecol. Evol.* **11**, 145–146 (1996).
- Mercot, H. & Poinot, D. *Entomol. Exp. et Appl.* (in the press).
- Lachaise, D. et al. *Evol. Biol.* **22**, 159–225 (1988).

Cranial surgery dates back to Mesolithic

Alt *et al.* have reported in Scientific Correspondence¹ what they thought to be the first 'unequivocal' evidence for the existence of healed trepanations (cranial surgery) from Ensisheim, Alsace, dated to 5100 BC. However, there is compelling evidence that such *intra vitem* surgery was carried out at an earlier date in eastern Europe, during the preceding Mesolithic period.

Until relatively recently, access to eastern European literature and archaeological material has been limited. Access has now become easier, and the Dnieper Rapids region of the Ukraine, about 400 km south-east of Kiev, yielded the evidence presented here. It comprises one aspect of a multi-disciplinary investigation into the chronology, diet and dental pathology of the human populations living in this area at the transition of the Mesolithic and Neolithic periods^{2,3}.

During analysis of the crania of about 307 skeletons from cemeteries dating back to each period, two individuals were found to exhibit evidence of trepanation. The cranium of the first individual (Fig. 1), a male who was more than 50 years old at death, has a healed lesion on the left side of the frontal bone, about 15 mm anterior to the coronal suture and 30 mm to the left of the sagittal suture.

The aperture has a pronounced raised border of remodelled bone, and 'stepping' in the central area reflects the progressive stages of closure during life. The central area of the depression is about 6 mm in diameter and the remodelled surface is extremely thin (less than 1mm) at this point. This skeleton (No. 6285-9) was recovered during excavations of the Vasilyevka II cemetery, carried out by A. D. Stolyar in 1953, and was originally reported in Russian in 1966⁴.

New accelerator radiocarbon dating of the Dnieper Rapids cemeteries^{5,6} has shown that the typological attribution of several of these cemeteries to the Neolithic period is erroneous. Three radiocarbon determinations of Vasilyevka II have placed this cemetery between 8,020 and 7,620 uncalibrated years before present, calibrated at the two sigma level to 7,300–6,220 BC (OxA-3804, OxA-3805 and OxA-3806).

Such dating confirms that this evidence offers the earliest example so far of trepanation, with the complete closure of the aperture indicating survival of the individual after surgery. The Vasilyevka II example predates the Ensisheim Neolithic site by 1,000–2,000 years. The remodelling that has occurred obscures the evidence for the method used to remove the bone, but it is

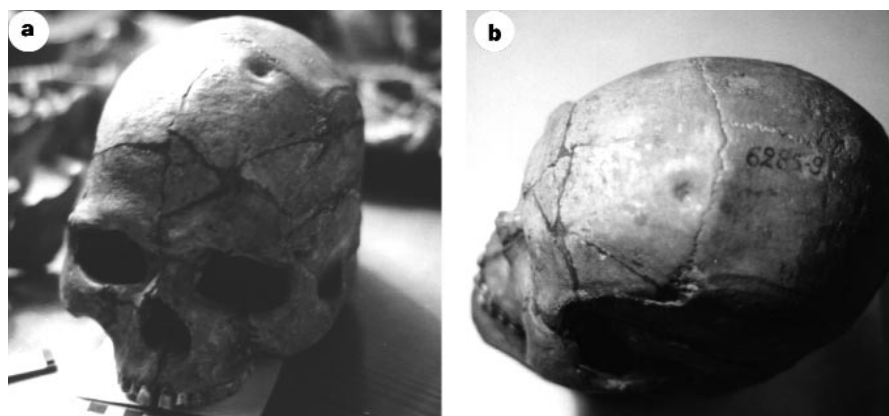


Figure 1 Photographs of cranium of skeleton No. 6285-9 from the cemetery of Vasilyevka II. **a**, front view: the healed trepanation is clearly visible as an area of remodelled bone on the left side. **b**, superior-lateral view: complete closure of the aperture and the 'stepped' nature of progressive stages of healing are apparent.

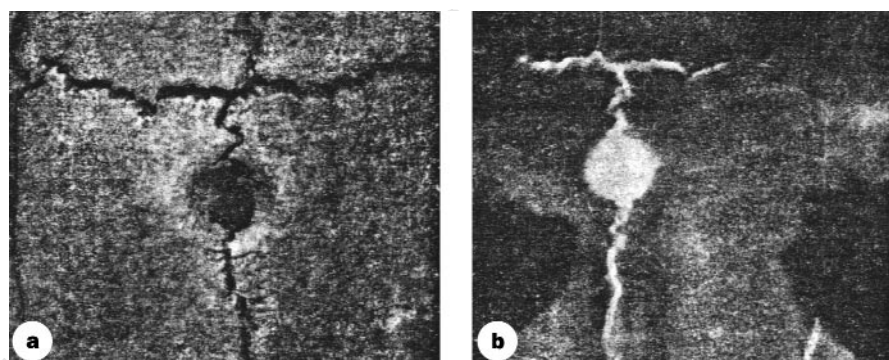


Figure 2 Skeleton No. 54 from the Vovnigi II cemetery (after Gokhman 1966:142). **a**, photograph of the unhealed trepanation, superior view. The front of the cranium is to the top of the picture. **b**, Radiograph of the trepanation highlighting the complete absence of remodelled bone.

likely that drilling⁷ would have been used, as was the case for the second individual found to exhibit evidence of trepanation.

This second trepanation was noted in Russian language texts in 1966 (ref. 4). An adult male (Fig. 2) from the Vovnigi cemetery complex (5,470–4,783 BC) exhibits an almost identical style of trepanation to that evident in Fig. 1. Individual No. 54 from the Vovnigi II cemetery exhibits an unhealed trepanation hole on the sagittal suture about 20 mm above the coronal suture. The bone defect on the external plate of the skull is of a regular circular shape (14 mm in diameter), while on the internal plate the hole is irregular, resembling a rhombus⁷. The diploë is completely closed around the aperture, and the edges of the internal plate are thin, exhibiting scratches made during drilling.

Although it is of a later date, this example confirms the validity of the earlier identification. In this second instance, however, the lesion does not exhibit remodelling to the same extent as that shown in Fig. 1a, suggesting that this individual did not survive the surgery for any great period of time.

The trepanations described by Alt *et al.*, although exceptional because of both their size and the survival of the patient after

surgery, cannot therefore be viewed as 'the earliest unequivocal evidence of healed trepanations yet discovered'¹. In the light of the above examples, it is possible that similar finds will emerge in other regions that have offered limited access to western researchers.

Accurate dating also needs to be considered. The recent accelerator radiocarbon dating of the Vasilyevka II cemetery suggests that archaeological cemeteries and sites dated by relative methods — such as typological seriation of artefacts — remain to be accurately dated in absolute terms.

Malcolm C. Lillie

Centre for Wetland Archaeology, University of Hull,
Hull, HU6 7RX, United Kingdom
e-mail: M.C.Lillie@geo.hull.ac.uk

- Alt, K. W. *et al.* *Nature* 387, 360 (1997).
- Lillie, M. C. *Curr. Anthropol.* 37, 135–142 (1996).
- Lillie, M. C. in *Invisible People and Processes: Writing Gender and Childhood into European Archaeology* (eds Moore, J. & Scott, E.) 213–228 (Leicester Univ. Press, London, 1997).
- Gokhman, I. I. *Naselenie Ukrainy v epokhu mezolita i neolita: antropologicheskii ocherk* (The Population of the Ukraine in the Mesolithic and Neolithic Periods: an Anthropological Outline) 141–142 (Nauka, Moscow, 1966).
- Potekhina, I. D. & Telegin, D. Ya. *Curr. Anthropol.* 36, 823–826 (1995).
- Jacobs, K. *Curr. Anthropol.* 34, 311–324 (1993).
- Lisowski, F. P. in *Diseases in Antiquity: a Survey of Diseases, Injuries and Surgery of Early Populations* (eds Brothwell, D. and Sandison, A. T.) 651–672 (C. C. Thomas, Illinois, 1967).

Celestial calculations

Pierre-Simon Laplace (1749–1827): A Life in Exact Science

by Charles Coulston Gillispie
Princeton University Press: 1988. Pp. 322.
\$49.50, £35

William R. Shea

Laplace began his major life-work, the mathematical exposition of Newton's law of universal gravitation, in 1773 when he was a young professor at the Ecole Militaire in Paris. It is usually said he was determined at all costs to vindicate Newton's theory against troublesome cases, such as why Jupiter's orbit appeared to be continuously shrinking while Saturn's continuously expanded. In the *Mécanique Céleste*, published in five volumes between 1799 and 1825, Laplace was to defend the invariability of planetary mean motions, and intimate that Newton had underestimated the power of his own theory when he concluded that divine intervention was necessary to preserve the Solar System in equilibrium.

But all this was in the future. In 1773, as Charles Coulston Gillispie discloses, Laplace was unable to make the two wayward planets fit in with the mathematical equations he derived from the mutual attraction of celestial bodies. He was not orthodox in his Newtonian beliefs, and he even tampered with universal gravitation by suggesting that gravity was propagated in time instead of instantaneously, so its intensity depended not only on the mass and the distance of bodies but also on their velocity.

How Laplace became the Newton of France is a tale of mathematical complexities that is admirably told by Gillispie. He revised for this book an earlier version that appeared as an entry in the monumental *Dictionary of Scientific Biography*, a work of 16 volumes that he edited between 1970 and 1980. The sequence, range and results of Laplace's work are presented against the background of the science of his day, and are explained in his own terms using his own notation.

Gillispie offers a scholarly analysis of Laplace's career, and shows that by 1786 he had largely resolved the vexed question of apparently cumulative discrepancies between the theoretical and observed positions of celestial bodies. Laplace began by establishing that the acceleration of Jupiter's mean motion and the deceleration of Saturn's were interdependent.

He then tackled the problem of the stability of the motions of Jupiter's satellites. This stability had been discovered by Galileo in 1610 and had enabled Ole Christensen Römer in 1676 to measure the speed of light by recording the dependence of the varia-



Laplace: the Newton of France.

tions in the apparent time of their eclipses on the distance between Jupiter and the Earth. The magnitude of the task faced by Laplace in working out the stability of the first three satellites can best be conveyed by mentioning that their motion depended on nine second-degree differential equations, the integrals of which contained 18 arbitrary constants, of which 12 could be determined from the eccentricities, orbital inclinations and positions of the nodes and aphelia, and the other six from the mean motions and their epochs. Laplace of course had no computer to help him, and he could not even rely on the accuracy of the data pertaining to most elements of the theory of Jupiter's satellites.

Laplace was assisted by a young mathematician and astronomer, Jean-Baptiste Delambre, who undertook the "delicate and painful task" — we might call it drudgery — of computing astronomical tables for Jupiter and Saturn. But it was Laplace himself who calculated the period of our own moon's secular variation in mean motion of revolution, a period so long that it amounted to millions of years.

Laplace reached the happy conclusion that the Moon would never come crashing down on us, and that one day it would actually retreat from the Earth. In 1796 he published a semipopular account of his celestial mechanics, the *Exposition du Système du Monde*, a model of French scientific prose and one of the most successful works of science ever composed. It included his famous 'nebular hypothesis' in which he suggested

that the Sun was originally surrounded by a gaseous atmosphere that contracted on cooling and formed the planets by condensations in the plane of the solar equator.

Laplace also wrote a 'popular' book, *Essai Philosophique sur les Probabilités*, published in 1814, that was intended as an introduction to the second edition of his comprehensive *Théorie Analytique des Probabilités*, in which he described the mathematical tools he invented to predict the probabilities that particular events will occur in nature. *Essai Philosophique* contains his celebrated claim that if enough is known about a physical system, its entire future can be predicted. For an intelligence that could comprehend, at a given instant, all the forces in nature and the respective situations of all the bodies in the Universe, "nothing", he wrote, "would be uncertain, and the future, as well as the past, would be present to its eyes". This idea was to prevail in classical physics until the introduction of quantum mechanics in the early part of the twentieth century.

Laplace became interested in experimental physics at the instigation of the chemist Antoine Lavoisier, who was studying the effects of temperature and pressure on the vaporization of liquids. Together they showed that respiration was combustion, and Laplace strengthened Lavoisier's hand in the battle against the idea of phlogiston by establishing that the hydrogen produced when an acid acts on a metal comes from the acid and not from the alleged phlogiston.

Laplace did not have strong political convictions and, unlike Lavoisier, who was beheaded in 1794, he does not seem to have been seriously inconvenienced by the French Revolution, except during the short Jacobin dictatorship when he thought it wiser to leave Paris. But he was an active member of the Bureau des Longitudes, and a founder of the Société d'Arcueil, named after the village south of Paris where he bought a property in 1806. Under Napoleon he held the post of Minister of the Interior for just six weeks and, after the restoration, he was made a marquis by Louis XVIII.

In writing this book, Gillispie was assisted by two leading historians of science, Ivor Grattan-Guinness and Robert Fox. Grattan-Guinness provides a concise but illuminating discussion of the 'Laplace transform', a method of solving differential and integral equations that is still widely used by mathematicians. Fox contributes four lucid chapters on work done in the Laplacian school on the velocity of sound, short-range forces, refraction and capillary action. He also shows why and how Laplace's physics was challenged by the development of Fresnel's wave theory of light and the discovery, by

Hans Christian Oersted in 1820, of the magnetic effect of passing of an electric current along a wire. This finding implied, contrary to Laplace's belief, that electricity and magnetism could interact with each other.

Gillispie's distinguished biography is a magisterial survey of one of the most influential scientists of the past two centuries. It is also a history of science at its most challenging. The 36-page critical bibliography is up to date and opens the way for future research by taking potential students a few steps along the right path. □

William R. Shea is in the Institut d'Histoire des Sciences, Université Louis Pasteur de Strasbourg, 7, rue de l'Université, 67000 Strasbourg, France. e-mail: william.shea@ifrest-ulp.u-strasbg.fr

Figure heads

The Number Sense: How the Mind Creates Mathematics

by Stanislas Dehaene

Oxford University Press: 1997. Pp. 274. \$25

Brian Butterworth

Numbers are one of the great foundation stones of our civilization. Without them, the vast and intricate edifices of technology, science and trade could never have been built. Albert Einstein called them "the symbolic counterpart of the universe".

Numerals have been used for almost as long as people have had agricultural surpluses to trade, at least 12,000 years. But where did the core concept of number — the idea of the 'many' of collections of things — come from? It is certainly not an easy idea. As Adam Smith wrote: "numbers are among the most abstract ideas which the human mind is capable of forming." Did some ancient Ramanujan, sitting perhaps in a mud hut in Ur, invent them, or is a sense of number part of the way we naturally categorize the world?

The answer lies in part in understanding how the human brain deals with numbers. What makes the problem interesting, and this book timely, is that studies of numerical abilities from developmental psychology to functional brain imaging are converging on the idea that humans have inherited brain circuits specialized for numbers. It is one of the great virtues and values of the book that it draws these findings together succinctly and with verve.

Stanislas Dehaene reviews the numerical abilities of different species of animals, as revealed in ingenious laboratory experiments. There are 'rat accountants' that can count the correct number of shocks before responding; a clever parrot that can say the number of red objects on a tray; and primates that can add two pairs of numbers and then select the larger sum. There is no doubt that such behaviour is controlled by the

numerosity of the stimuli and not by some covarying factor. But why should animals possess numerical abilities? To understand this, we need evidence, rather than speculation, about how animals use numerosity in the wild, not in the laboratory.

There are human infants as young as one day old who can distinguish visual arrays on the basis of numbers up to about four, and a few months later can add one and one, and subtract one from two. For them to do this, Dehaene points out, the ability must be innate, as the babies have not had time to learn. Dehaene is therefore arguing that the influential Swiss psychologist Jean Piaget was fundamentally wrong. Piaget believed that his experiments showed that children build the concept of number out of other, more basic, logical concepts and processes, such as transitive inference, one-to-one correspondence and an ability to abstract away from the sensory properties of the things being counted. The role of numerals and counting words was marginalized.

Dehaene is clearly angry about the influence that both Piagetians and the formalist mathematicians from the Bourbaki group have had on French education. "The reformers thought that children should be familiar with the general theoretical principles of numeration before being taught the specifics of the base-10 system. Hence, believe it or not, some arithmetic textbooks started off by explaining that $3+4$ is 12 — in base 5! It is hard to think of a better way of befuddling a child's thinking."

Work on patients with brain lesions, including Dehaene's own patients, shows that there are special areas of the brain, probably the left and right inferior parietal lobes, that could contain the inherited number circuits. They seem to carry out very basic numerical processes such as comparing the magnitude of two numbers. The left parietal lobe alone is responsible for more advanced functions, and when it is damaged patients can suffer from very specific numerical deficits, such as being unable to remember multiplication facts, despite being able to remember non-numerical facts. They may even be able to carry out other mathematical operations, including addition and subtraction. There are even reports of patients who can no longer do simple arithmetic but can still do algebra.

Dehaene thinks that these circuits do not represent numerosities as such but continuous approximations to them that he calls "analogue magnitudes". The basic mechanism, "the accumulator", parses the world into discrete units and stores in a memory an approximately equal quantity for each unit. The total accumulated quantity then represents the number of units detected. For Dehaene, therefore, the number sense is a kind of quantity sense. To get from these fuzzy representations to the discreteness of integers requires the uniquely human ability

to create and use symbols, including number words, numerals and the idea of numerosities.

But why is the brain unable to represent discrete numerosities in the first place, as perceptual systems are designed to isolate discrete objects in the world? Indeed Dehaene, both in the book and elsewhere, is sometimes happy to talk of "numerosity detectors" that create discrete representations from stimulus information. These processes, called "subitizing", enable even infants to distinguish at a glance, without counting, numerosities up to about four. Whether the brain represents numerosities as discrete or continuous quantities is an important issue, and its resolution will have consequences for our ideas of both brain specialization and the evolutionary history of the number sense.

Our ability to use numbers is as much a fundamental aspect of what makes us human as is our ability to use language. This book provides an excellent introduction to how this ability is organized in our mathematical brain. □

Brian Butterworth is in the Department of Psychology and the Institute of Cognitive Neuroscience, University College London, Gower Street, London WC1E 6BT, UK.

ADVERTISEMENT

Check out this WEB SITE!

Reviews on Cancer Online

<http://www.elsevier.com/locate/roco>
<http://www.elsevier.nl/locate/roco>

Old embryos and new developments

Principles of Development

by Lewis Wolpert with Rosa Beddington, Jeremy Brookes and Thomas Jessell
Current Biology/Oxford University Press:
1998. Pp. 484. £25.95, \$65

Embryology: Constructing the Organism

edited by Scott F. Gilbert and Anne M. Raunio
Sinauer: 1997. Pp. 240. \$69.95, £32.95(hbk), £12.99 (pbk)

Jonathan Slack

Twenty years ago it was possible to pick up two textbooks of developmental biology written by different authors and find almost no common ground in the contents. That was because, as recently as the 1970s, there were profound disagreements about what the basic questions of development actually were, and what sort of things we might accept as a satisfactory explanation.

To some, it was all about the nuts and bolts of the control of gene expression, to others it was all about cell movement, and there was even a small group to whom it was really a matter of guarding the true faith of classical experimental embryology, in the hope that future generations would have the means to solve the problems.

But those days are gone and the basic problems are now solved, at least in principle. Today, when we pick up a textbook and flip through it, we find pretty much what we expect to find. This consensus means that the subject has reached maturity. In the past 20 years we have understood things that were previously mysterious, and we understand them well enough to have some confidence in relating the findings to undergraduates.

Lewis Wolpert and his co-authors have been at the forefront of the revolution that has brought the subject so far. *Principles of Development* is an ambitious attempt to capture the world undergraduate market from the main competition, Scott Gilbert's *Developmental Biology*, now in its fifth edition. In this regard he has the advantage of starting afresh, whereas Gilbert's book has evolved from a first edition written when the subject was in a far less secure state. As someone who has recently had to design some undergraduate courses, I think Wolpert has done well, both in relation to the selection of which topics to include and in the level of detail at which they are presented.

The book starts by introducing the model organisms on which most developmental biology research is done: *Xenopus*, the chick, the mouse, the zebrafish, *Drosophila*, *Caenorhabditis elegans* and *Arabidopsis*. The next few chapters cover the formation of the



Embryologist in the making? A six-week human embryo.

basic body plan and include topics such as mesoderm induction, somitogenesis and *Drosophila* developmental genetics. Then there is a quick gallop through selected topics in a variety of non-*Drosophila* invertebrates and plants, before a return to vertebrates for various subjects including morphogenetics movements, cell differentiation, sex determination and neuronal development. I was surprised to find that "organogenesis" consists almost entirely of the limb and *Drosophila* discs, as there are, after all, a lot of organs in the vertebrate body apart from the limb. As usual, the last chapter deals with the evolution of developmental mechanisms, known in the trade as "Evo-Devo".

The book uses a consistent style of coloured illustration throughout that is the hallmark of the co-publisher Current Biology. These figures are mainly very clear and attractive, although in a few places the number and brightness of colours can impede appreciation of the content.

Students often find developmental biology difficult. This is partly because they now tend to have a poor knowledge of animal structure, so the necessary minimum of descriptive morphology is all totally new to them. It is also because developmental biology is essentially a synthesis of anatomy, molecular biology and genetics. Students are used to learning one thing at a time, and it can be confusing to jump from a genetic cross to a signal transduction pathway to an altered anatomy all in a few minutes.

Wolpert has taken great care to make things as clear and simple as he can but the subject is still challenging. To take a random example: "As a result of Toll receptor activation cactus protein is degraded and no longer binds the dorsal protein, which is then free to enter the nuclei. In embryos lacking cactus protein almost all the dorsal protein is found in the nuclei; there is a very poor concentration gradient and the embryos are ventralized." A research developmental biologist will absorb such a sentence immediately, but a student busy with other subjects from receptor biochemistry to animal behaviour needs to have it unpacked slowly and

methodically. A way to keep things simple is to stick to one organism at a time. This book contains several chapters that jump around between the seven organisms and require the reader to have instant recall of many names of genes and body parts. For this reason it may not be quite as clear as it first seems.

Gilbert, author of the present standard textbook on developmental biology, has found time, with Anne Raunio, to edit a completely different kind of book: *Embryology: Constructing the Organism*. At first sight the title might seem a needless risk, as everyone knows that the word "embryology" sounds old-fashioned to American ears, and in today's cut-throat world no academic book can afford to seem old-fashioned. But Gilbert has clearly planned this as a resource for the rapidly growing field of Evo-Devo, because it is a phylum-by-phylum account of the descriptive embryology of animals.

'Descriptive embryology' is what people used to do in the nineteenth century, before 'experimental embryology' was invented, and it means just describing what you can see happening down the microscope. This may sound very boring, but in fact it grows on you, and nobody can possibly understand anything about the development of an organism without a firm grasp of its descriptive embryology. In the past, anyone who wanted to know about this in an animal other than the six models had to find a copy of Kumé and Dan's *Invertebrate Embryology* (1957), which was almost unobtainable until it was reprinted by Garland in 1988, and is once again out of print.

Gilbert's book covers all the major animal phyla and also some of the minor ones for which the available information is hardly sufficient for a whole chapter (such as the lophophorate phyla). The approach is fiercely comparative, so the chapter about insects does not dwell too long on *Drosophila*, and that about mammals finds space for monotremes and marsupials as well as the mouse. There is even a token chapter on plants, which does not confine itself to *Arabidopsis* but mentions bryophytes and pteridophytes as well.

Although entirely in black and white, the illustrations are excellent and make this an attractive book. Because of its vast scope it contains a lot of very interesting stuff. I am always impressed by our ignorance of the mechanisms of asexual reproduction by budding, as enjoyed by many cnidarians, bryozoans and tunicates. This is a problem unsolved even in principle. I am also amazed by my own ignorance of so many things known to other people, but thanks to Gilbert I am now at least acquainted with the "groping penis" of the barnacle.

I am struck by the fact that, despite the relative obscurity of its contents, the book is fairly easy reading compared with Wolpert's, as there is no need here to understand the

complex genetic pathways involving repression of repressors. Students who find modern developmental biology difficult would find it a lot easier if they had mastered this material first. So logically they should be required to take a course in descriptive animal embryology using Gilbert's book, and then progress to developmental biology I and II using Wolpert's.

However, the appetite for animal structure in most places is so small that I suspect this could happen only in universities with a very strong zoology programme. In most places Gilbert's book is likely to be relegated to small but eager graduate classes studying Evo-Devo. Still, as the authors claim in the preface, this is a "timeless" book and it will still be useful when fashions change. □

Jonathan Slack is in the Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

The great elusionist

The Elusive Neutrino: A Subatomic Detective Story

by Nickolas Solomey

W. H. Freeman/Scientific American Library: 1997. Pp. 206. \$34.95, £19.95

Stephen Battersby

The manner in which a nothing-particle was first reluctantly proposed and then triumphantly displayed is one of the truly exciting adventures of science.

This is from Isaac Asimov's *The Neutrino* (1966), an early predecessor of Nickolas Solomey's new book, *The Elusive Neutrino*. Asimov does tell an exciting story, speeding through much of the history of science to explain what conservation laws are and why they are important. Only then does he describe how three such laws — the conservation of energy, momentum and angular momentum — led Wolfgang Pauli to postulate the existence of the neutrino in 1931.

The biggest problem of the day in particle physics was that the energy and momentum sums in beta decay (a form of radioactivity) didn't add up. Pauli realized that the problem could be solved by assuming that a new particle was escaping undetected from experiments, carrying away energy, momentum and angular momentum, but with no charge and little or no mass. The neutrino (little neutral one) came to be a vital part of particle physics, and eventually any lingering discomfort at this 'cheat' was removed when physicists detected the particle directly. That detection paved the way for neutrino astronomy, which was just beginning in 1966.

In the 30 years since Asimov's book, a revolution in fundamental physics has taken place. In the Standard Model of particle physics, three types of neutrino, three types of electron

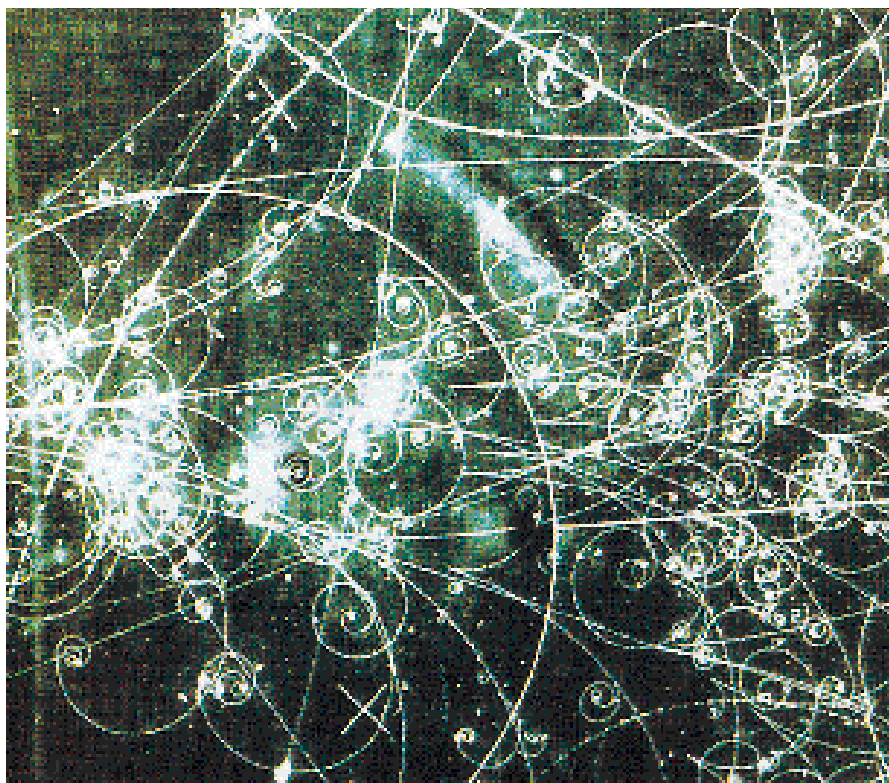


Image-makers: tracks left by the many different particles produced in a neutrino reaction.

and six types of quark interact by exchanging various force-carrying particles. And neutrino astronomy has produced results: a puzzling deficit of observed neutrinos from the Sun is well established, and, as Asimov hoped, neutrinos have been seen from a supernova (in 1987).

In the past couple of years, there have even been hints that neutrinos have mass, a property that could explain the solar neutrino deficit and the large-scale structure of the Universe and would constitute a step beyond the Standard Model.

In *The Elusive Neutrino*, Solomey's goals are necessarily greater than Asimov's because of these huge advances in our understanding. In any case, Asimov has few peers as a science writer, so it is unfair to compare the new book with the earlier. Unfair, but irresistible.

More than halfway through *The Neutrino* comes chapter seven, "Enter the neutrino". At that point in the proofs, Asimov's editor wrote "At last!". But that's the way it stayed. Solomey's book might have benefited from similar restraint. Asimov knew that his readers needed educating before they could understand the significance of the neutrino, as shown by his chapter titles: "Momentum", "Energy", "Atomic structure", "Mass-energy", "Electric charge". Solomey, instead, skates over the basics in chapter one, getting all the way to Fermi's four-point theory of beta decay. Then in chapter two he plunges straight into one of the more speculative topics in the book, "Cosmology and the neutrino mass". Perhaps Solomey thought a long introduction would bore the reader.

Another problem is that, like beta decay, parts of the book do not add up: vital particles of meaning seem to be escaping undetectably. Explanations are routinely let down by the language. It is not just a matter of distracting infelicities ("somewhat close proximity", "not as fully complete", "was impinging on the fluorescent screen"); it is often language careless enough to confuse or misinform the lay reader.

In the introduction, Solomey mentions the electron, proton and neutron, and then "particles that are even more elementary: quarks and leptons". The electron is a lepton, so here it is rendered more elementary than itself. Time reversal for macroscopic objects is "forbidden outright" (to contrast with "might not be plausible" for particles) and then in the next paragraph becomes "not a plausible occurrence". Later we read that "Because [random number generators] never repeat a sequence of numbers, they are as unpredictable as a roulette wheel". That is just nonsense.

The book is wide-ranging and, for the most part, informative — enough to make it fairly interesting. The only parts I found dull might reasonably interest an experimental physicist: Solomey is meticulous in apportioning credit for the various discoveries, and he describes scores of experiments in detail. He also discusses many of the experiments that got it wrong, or were not sensitive enough to be definitive. Surely two or three of these would have been enough to make the point that science is a messy process? □

Stephen Battersby is an assistant editor at Nature.

Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7

Jie-Oh Lee*, Alicia A. Russo & Nikola P. Pavletich*

Cellular Biochemistry and Biophysics Program and *Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The pocket domain of the retinoblastoma (Rb) tumour suppressor is central to Rb function, and is frequently inactivated by the binding of the human papilloma virus E7 oncoprotein in cervical cancer. The crystal structure of the Rb pocket bound to a nine-residue E7 peptide containing the LxCxE motif, shared by other Rb-binding viral and cellular proteins, shows that the LxCxE peptide binds a highly conserved groove on the B-box portion of the pocket; the A-box portion appears to be required for the stable folding of the B box. Also highly conserved is the extensive A–B interface, suggesting that it may be an additional protein-binding site. The A and B boxes each contain the cyclin-fold structural motif, with the LxCxE-binding site on the B-box cyclin fold being similar to a Cdk2-binding site of cyclin A and to a TBP-binding site of TFIIB.

Inactivation of the retinoblastoma (Rb) tumour suppressor is a common event in diverse human cancers (reviewed in ref. 1), and can occur through genetic alterations in Rb^{2–5}, or in the cellular p16^{MTS1}, cyclinD or Cdk4 proteins that regulate Rb¹. In cervical cancer, the inactivation of Rb is often through the binding of the E7 protein of the human papilloma virus (HPV)⁶, which is associated with over 90% of these cancer cases (reviewed in ref. 7).

Introduction of Rb into Rb-negative tumour cell lines can lead to growth arrest in culture and to the loss of tumorigenicity in nude mice^{8–10}. Rb exerts these growth-inhibitory effects in the mid-to-late G1 stage of the cell cycle^{11,12}, and seems to control the cell's commitment to DNA replication. The growth-suppressive activity of Rb is tightly coupled to the cell cycle, being regulated by cyclin-dependent kinase-mediated phosphorylation¹². Underphosphorylated, active forms of Rb dominate in G1, whereas hyperphosphorylated, inactive forms appear as cells enter S phase and dominate in G2 and M phases¹³.

There is increasing evidence that the function of Rb that is central to its growth-inhibitory effects is its regulation of the E2F family of transcription factors, which are required for the expression of genes involved in the G1–S transition and DNA replication^{14,15}. When underphosphorylated, Rb can bind to E2F, inhibit its transactivation function^{16–18} and, at certain promoters, can convert E2F to an active repressor of transcription^{19,20}. The inactivation of Rb by hyperphosphorylation leads to the release of transactivation-competent E2F.

The HPV E7 oncoprotein^{6,7}, which can cooperate with Ras to transform primary cells and induce tumours in transgenic mice²¹, inactivates Rb, at least in part, by causing the release of active E2F^{6,22}. Binding to Rb is essential for the ability of E7 to transform cells, because E7 proteins from HPV strains associated with a low risk for cervical cancer show reduced or no affinity for Rb²³. A similar Rb-inactivating function is shared by the simian virus 40 large T antigen²⁴ and the adenovirus E1A proteins²⁵. The E7, T antigen, and E1A oncoproteins contain a region of homology with the core sequence LxCxE that is necessary and sufficient for Rb binding, although other regions are also required for the disruption of the Rb–E2F complex and for cell transformation^{6,23,26}. The LxCxE sequence motif is found in certain cellular Rb-binding proteins, such as the D-type cyclins¹, but not in the E2F transcription factors, which share a distinct 18-residue Rb-binding motif^{14,27}.

Several lines of evidence indicate that the portion of Rb that is central to its biological effects is the ~45K 'pocket' region, which

consists of the A and B boxes that are conserved in the Rb-related p107 and p130 proteins. The pocket is where most of the Rb-binding cellular and viral proteins make their primary contacts^{28–30,14,15}; it is necessary and sufficient for transcriptional repression when fused to a heterologous DNA binding domain^{19,20}; and it is the primary target of genetic alterations observed in tumours^{28–30}. The Rb pocket is also necessary for growth inhibition, although it is not sufficient because additional sequences from the ~15K C-terminal region of Rb are also required^{10,16}.

Here we describe the crystal structure of the Rb pocket bound to an HPV-16 E7 LxCxE peptide, and discuss the implications of this structure for understanding the protein-binding activity of Rb, its regulation, and its inactivation in cancer.

Structure determination

The minimal regions of Rb and HPV-16 E7 required for tight complex formation are the ~45K pocket of Rb^{28–30} and a nine-residue peptide³¹ from the conserved region 2 (CR2) of E7 that contains the LxCxE sequence²³. This E7 peptide binds the Rb pocket with a dissociation constant (K_d) of 110 nM (Table 1) and, according to competition experiments³¹, has a 20-fold weaker Rb affinity compared with full-length E7 protein.

To facilitate crystallization, and based on studies showing that the A and B boxes can associate noncovalently³², a 46-amino-acid portion of the spacer in between the A and B boxes of the Rb pocket was removed by digestion at engineered flanking thrombin sites (Fig. 1a). The spacer, which is the least conserved region of Rb (Fig. 1a), and is different in the related p107 and p130 pocket proteins, appears to be unstructured or loosely folded based on proteolytic digestion (data not shown). The removal of the spacer

Table 1 Dissociation constants (nM) of Rb complexes with E7 or E2F1 peptides

	A-spacer-B	A + B	GST-A-spacer-B-C
CR2 of HPV16 E7	110 ± 34	110 ± 15	210 ± 116
E2F1 peptide	400 ± 82	490 ± 76	1090 ± 260
E2F1 peptide after saturation with CR2	510 ± 82	590 ± 49	ND

A-spacer-B is residues 379–792; A + B is the crystallized pocket domain, residues 372–589 + 636–787; GST-A-spacer-B-C is residues 364–928 of Rb fused to GST. CR2 of HPV16 E7 is AcO-DLYCYEQLN-CONH₂; E2F1 peptide is AcO-LDYHFGLEEGEGIRDLFD-CONH₂. A p53 transactivation domain peptide (SQETFSDLWKLLPEN) as a control gave no detectable signal. The error range is that resulting from fitting the data to the two-state binding model. ND, not determined. The large error ranges associated with measurements with GST-A-spacer-B are presumed to be due to this protein's tendency to aggregate.

did not detectably affect the structural integrity or the protein-binding activity of the pocket: the A and B boxes remained quantitatively associated as determined by gel filtration chromatography (data not shown), and the pocket bound the E7 LxCxE peptide or the 18-residue Rb-binding motif of E2F with the same affinities as

the native pocket, as determined by isothermal titration calorimetry (Table 1).

The crystal structure of this engineered ~41K Rb pocket (residues 372–589 and 635–787) bound to the nine-residue E7 LxCxE peptide was determined with the multiwavelength anomalous

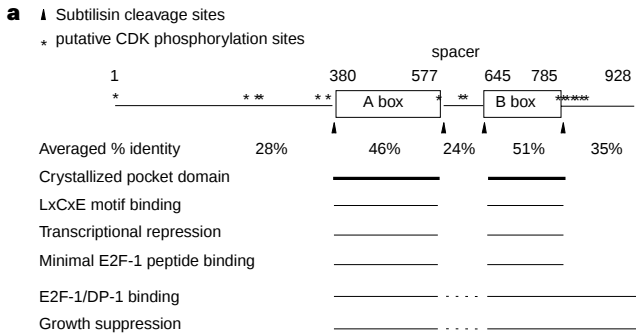


Figure 1 The A–B pocket is central to Rb function. **a**, Functions of Rb discussed in the text and the regions that underlie these functions are summarized; the dashed line for the spacer in the last two entries indicates that its requirement for those functions has not been studied. The averaged identity is that among the Rb homologues from human, newt, chicken, fruit fly and maize, and among the related human p107 and p130 proteins. The function of the ~40K N-terminal region preceding the pocket is not clear, although it is not required for growth suppression and appears to not participate in binding the cellular or viral proteins. This N-terminal region is also largely absent from the maize homologue of Rb. In the Rb-related p107 and p130 proteins, the spacer region is larger by ~15K and has no apparent similarity to that of Rb. The putative Cdk phosphorylation sites are those with the S/T-P sequence. **b**, Experimental MAD electron density of the LxCxE peptide calculated at 2.5 Å resolution and contoured at 1.0 σ , with the 1.85 Å refined model superimposed. Because of the binding of a mercury atom to the cysteine of the peptide, some of the E7 MAD electron density is shifted with respect to the model that is refined against native data. **c**, Overall view of the Rb pocket–LxCxE peptide complex structure. The dashed lines indicate the regions that are disordered in the structure.

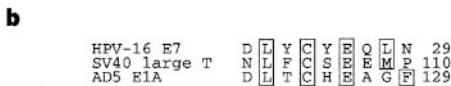
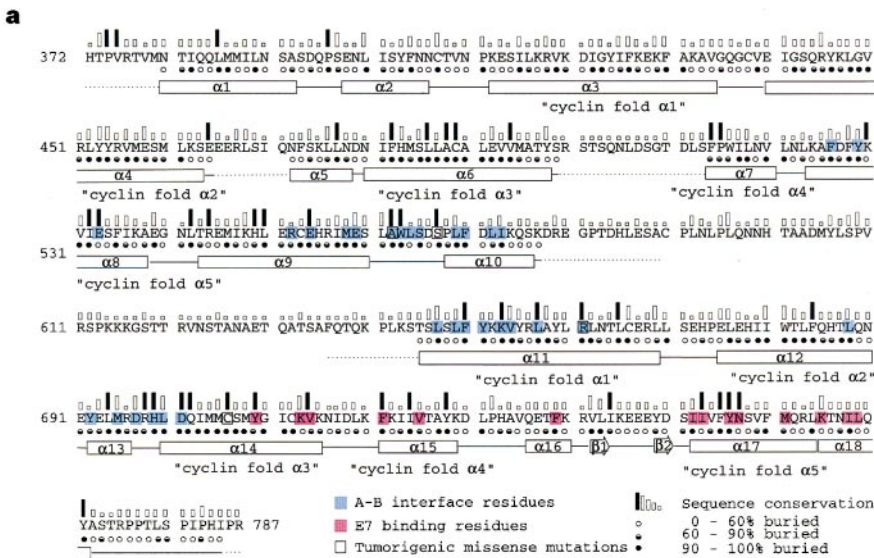
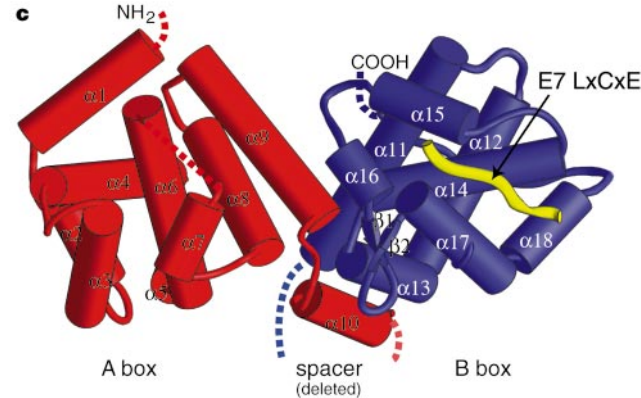
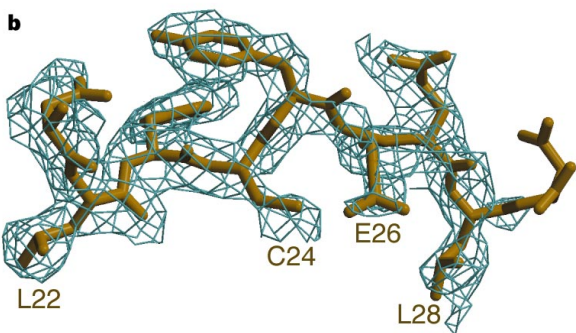


Figure 2 Conserved residues of the Rb pocket map to the LxCxE binding site or to the A–B interface, as well as to the hydrophobic cores of the A and B boxes. **a**, The bars indicating sequence conservation are black for residues that are identical in the Rb homologues from five species: human, newt, chicken, fruit fly and maize, and decrease in height for lower levels of identity. The dashed lines in the secondary structure assignments indicate residues that are present in the crystallized pocket but are disordered in the crystals. The helices of the A and B box that correspond to the five helices of the cyclin fold are labelled. **b**, The HPV16 E7 peptide used in the co-crystallization aligned with the functionally homologous regions of the T antigen and E1A oncogenes.

diffraction (MAD) method using a mercury derivative, and was refined at 1.85 Å resolution to a crystallographic *R*-factor of 21.8% (Fig. 1b, Table 2).

Overall structure of the complex

As was observed with the crystal structure of the isolated Rb A box³³, and predicted for the B box³⁴, the A and B boxes each contain the cyclin fold, a five-helix structural motif also present in the structures of cyclins³⁵ and the basal transcription factor TFIIB³⁶. In addition to the 10 helices of the two cyclin folds, the Rb pocket contains eight other helices, a β -hairpin, and an extended tail (Figs 1c, 2a), with these additional structural elements being more numerous than those that augment the cyclin folds in the structures of cyclins or TFIIB.

The E7 LxCxE peptide binds in an extended conformation to a shallow groove on the B box (Fig. 1b). Intermolecular contacts are made by the alternating Leu 22, Cys 24, Glu 26 and Leu 28 side chains, which point into the B-box groove, and four backbone groups of the E7 peptide (Fig. 2b).

The interface between the A and B boxes is extensive and can be considered as an integral part of the overall pocket structure. The A–B interface has a hydrophobic core that is continuous with that of the B box, suggesting that the B box needs the A–B interface for its stable folding.

The surface residues of Rb that are conserved across species and in

the p107 and p130 family members cluster on two regions (Fig. 2a). One is the LxCxE binding site on the B box, indicating that this binding site is central to the cellular function of Rb. The second conserved region is on the A–B interface, suggesting that it may participate in binding to E2F or to proteins that may mediate transcriptional repression by Rb. The missense mutations identified in tumours map to the A–B interface or the B box (Fig. 2a), supporting the significance of these conserved regions.

A comparison of this Rb–E7 complex with the cyclin fold containing cyclin A–Cdk2 (ref. 35) and TFIIB–TBP–DNA³⁶ complexes reveals that the LxCxE binding site on the Rb B box closely corresponds to the primary Cdk2-binding site of cyclin A, and to a TBP-binding site of TFIIB. The corresponding region of the A box is involved in forming the A–B interface.

The Rb pocket

Structures of the A and B boxes. The cyclin fold, which forms the basis of the A- and B-box structures, consists of a three-helix bundle with two additional helices packing on two sides, giving a five-helix core motif ($\alpha 3$, $\alpha 4$, $\alpha 6$, $\alpha 7$, $\alpha 8$ for the A box, and $\alpha 11$, $\alpha 12$, $\alpha 14$, $\alpha 15$, $\alpha 17$ for the B box; Fig. 1c). The additional eight helices, β -hairpin, and extended tail of the Rb pocket serve either at the A–B interface ($\alpha 9$, $\alpha 10$, $\alpha 13$), or the LxCxE binding site ($\alpha 16$, β hairpin, $\alpha 18$), or cover surface regions of the A and B cyclin folds ($\alpha 1$, $\alpha 2$, $\alpha 5$, tail).

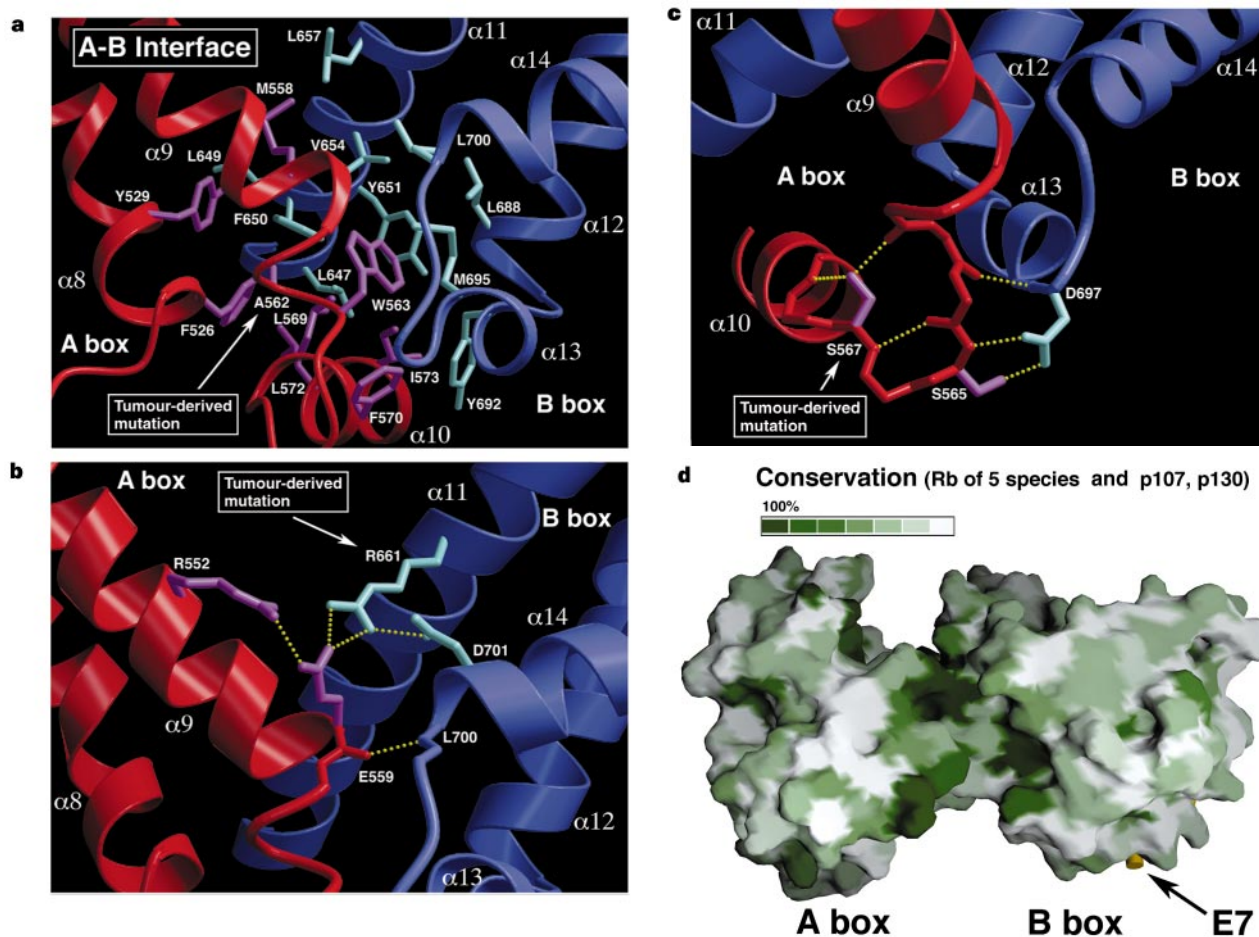


Figure 3 The A–B interface is like a folded structural domain, and is highly conserved. **a**, Residues that form the hydrophobic core of the interface. **b**, **c**, Two regions with notable hydrogen-bond networks involving both side-chain and backbone groups. Hydrogen bonds are indicated by yellow dotted lines. The importance of the interface is underscored by the presence of tumour-derived mutations that target residues either in the hydrophobic core or in the hydrogen-

bond networks. **d**, The surface of the pocket is coloured according to residue identity across the Rb homologues of five species and the p107 and p130 family members (Fig. 2a). Dark green indicates identities and white non-conserved residues. This view highlights the striking conservation of residues that are on a surface patch about halfway around the A–B interface.

Structure of the A–B interface. The A–B interface forms through the packing of portions of seven helices, three from the A box ($\alpha 8$, $\alpha 9$ and $\alpha 10$) and four from the B box ($\alpha 11$, $\alpha 12$, $\alpha 13$, $\alpha 14$), and has a buried surface area of $2,013 \text{ \AA}^2$ (Figs 1c, 3a). A distinguishing topological feature of the interface is the intermingling of A-box and B-box helices (Fig. 3a). The interactions between the A and B boxes are mediated by a compact hydrophobic core of about 20 side chains (Fig. 3a) and networks of backbone and side-chain hydrogen bonds (Fig. 3b, c). In these respects, the packing at the interface resembles the folding of a small globular protein.

The contributions of the A and B boxes to this interface structure differ fundamentally. The interface elements of the A box are mostly external to its cyclin fold, and the A-box hydrophobic core does not participate in the interface. In contrast, the interface elements of the B box are mostly integral to its cyclin fold, and the hydrophobic core of the B box is continuous with that of the interface. This suggests that the B box requires the interface, and thus indirectly the A box, for its folding. This model is supported by differential scanning calorimetry, where the Rb pocket unfolds in a single cooperative transition (data not shown), and may explain why the A box is needed for LxCxE binding^{28–30}, even though this motif only interacts with the B box.

The importance of the interface structure for Rb function is indicated by the overwhelming conservation of interface residues (Figs 2a, 3d), and by the targeting of interface residues by tumori-

genic missense mutations^{2–5}, both in the hydrophobic core (Ala 562; Fig. 3a), and in the hydrogen-bond networks (Arg 661, Ser 567, Fig. 3b, c).

The Rb–E7 complex

B box–LxCxE contacts. The nine-residue LxCxE peptide binds a shallow groove on the B box in an extended, β -strand-like conformation (Fig. 4a). The alternating Leu 22, Cys 24, Glu 26 and Leu 28 side chains point into the B-box groove and make intermolecular contacts, whereas the remaining residues (Asp 21, Tyr 23, Tyr 25, Gln 27 and Asn 29) point out towards the solvent and do not participate directly in binding (Fig. 4a). The structural framework of the shallow groove is formed by three cyclin-fold helices, with $\alpha 17$ forming one side and the $\alpha 14$ and $\alpha 15$ helices forming the other side of the groove. The bottom of the groove is lined with residues from the hydrophobic core of the B-box cyclin fold, whereas the ends are capped by the $\alpha 16$ and $\alpha 18$ non-cyclin-fold helices (Fig. 4a).

There is a high density of van der Waals and hydrogen-bond intermolecular contacts distributed uniformly along the central seven-residue portion of the LxCxE peptide. At the ends, the Leu 22 and Leu 28, and at the centre, the Cys 24 side chains of the peptide bind in hydrophobic pockets in the groove (Fig. 4a, b). Hydrogen bonds are made by three backbone carbonyl groups (residues 22, 23 and 24), by one backbone amide group (residue 24), and by the Glu 26 carboxyl group of the peptide (Fig. 4a). Complex formation buries a total of $1,150 \text{ \AA}^2$ of surface area.

Conservation of the LxCxE motif. The observed binding explains the conservation of the LxCxE motif in the viral oncogenes (Fig. 2b).

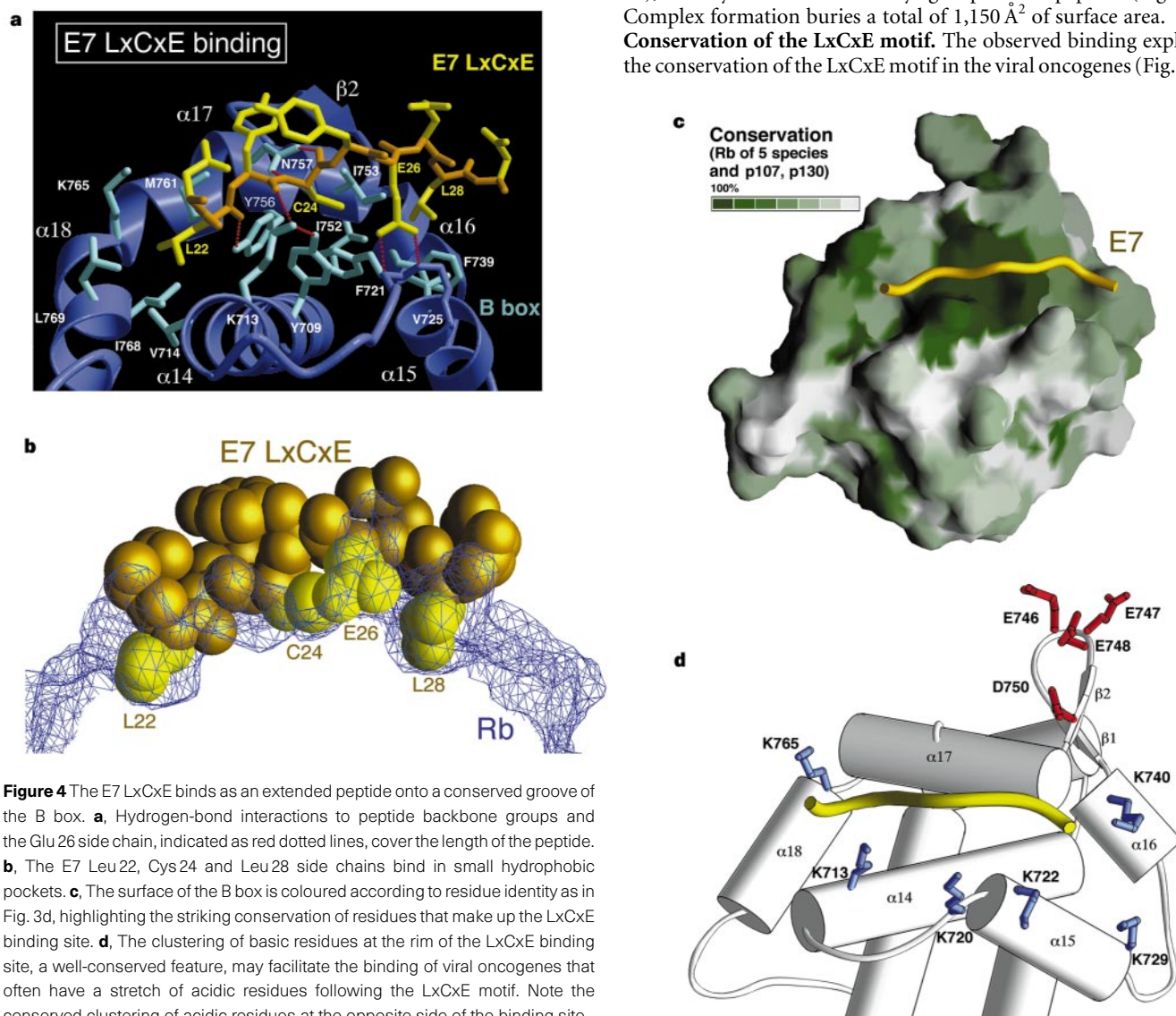


Figure 4 The E7 LxCxE binds as an extended peptide onto a conserved groove of the B box. **a**, Hydrogen-bond interactions to peptide backbone groups and the Glu 26 side chain, indicated as red dotted lines, cover the length of the peptide. **b**, The E7 Leu22, Cys24 and Leu28 side chains bind in small hydrophobic pockets. **c**, The surface of the B box is coloured according to residue identity as in Fig. 3d, highlighting the striking conservation of residues that make up the LxCxE binding site. **d**, The clustering of basic residues at the rim of the LxCxE binding site, a well-conserved feature, may facilitate the binding of viral oncogenes that often have a stretch of acidic residues following the LxCxE motif. Note the conserved clustering of acidic residues at the opposite side of the binding site.

Leu 22 and Cys 24 (bold in LxCxE) bind hydrophobic pockets with tight complementarity (Fig. 4b). As the peptide backbone in this region is bound by four hydrogen bonds, there is little freedom for the peptide to move to accommodate binding by other side chains. The preference for a Cys 24 rather than the isosteric Ser (ref. 31) is consistent with the overall hydrophobic nature of the pocket where this side chain binds (Fig. 4a). The carboxylate group of Glu 26 (bold in LxCxE) makes two hydrogen bonds to two backbone amide groups of the α 15 helix, the strength of these hydrogen bonds being augmented by the positive dipole moment at the amino terminus of the helix (Fig. 4a). Accordingly, the Glu 26Gln mutation reduces binding by over a factor of 100 (ref. 31). The Leu 28 (bold in LxCxE) binding pocket at one end of the groove, although favourably hydrophobic, is broader (Fig. 4b). This suggests that, although this residue is needed for tight binding, there may be some flexibility as to its size and position. Accordingly, the T antigen has a Met, and E1A has a Phe after a glycine insertion, at the positions corresponding to Leu 28 of E7 (Fig. 2b).

The alternating LxCxE residues that point away from the binding site seem to have minor roles in stabilizing the extended conformation of the peptide: Tyr 23 and Tyr 25 stack with each other, and Asp 21 makes a long-distance hydrogen bond to Tyr 25 (Fig. 4a).

Conservation of the LxCxE-binding site. The structure reveals an overwhelming conservation of residues that make up the B-box site (Fig. 4c). Most important is the conservation of the four residues that contact the backbone of the peptide (Figs 2a, 4a): Tyr 709, Tyr 756 and Asn 757 are identical in all five Rb homologues from diverse species and in the p107 and p130 family members, whereas Lys 713 is identical in all but the maize Rb homologue, where it is an Arg (ref. 37). Also important is the invariant presence of an extra non-helical residue (residue 757) in the midst of the α 17 helix (Fig. 2a). This one-residue insertion gives the α 17 helix a kink that is essential to the curvature of the binding site (Fig. 4a).

The conservation of this B-box site indicates that it plays a central role in Rb function, and that this role is common to the p107 and p130 family members³⁸. It is likely that the D-type cyclins, which can bind to Rb and have an LxCxE motif conserved from humans to plants³⁷, contact this site. It is also likely that binding to other cellular proteins with an important role in the Rb pathway might have contributed to the high conservation of this site.

An unusual feature of the LxCxE binding site is a separation of a substantial amount of charge (Fig. 4d). This charge separation is unlikely to be coincidental because the residues responsible for it are conserved overall and most do not have apparent structural roles (Fig. 2a). The concentration of Lys residues at the rim of the binding site may explain why the viral oncogenes have all evolved a stretch of 3–6 acidic residues immediately after the LxCxE motif²³, as the

attraction between the oppositely charged residues may facilitate their association.

Pocket alterations in cancer

Most of the Rb alterations identified in tumours involve gross changes such as nonsense, frameshift, deletion and splice-site mutations that affect the pocket^{2–5}. The less frequent missense mutations that have been described map to conserved residues at the A–B interface or the B box. At the A–B interface, Ala 562 (ref. 39) maps to the hydrophobic core (Fig. 3a), whereas Arg 661Trp (ref. 5) (Fig. 3b) and Ser 567Leu (ref. 2) (Fig. 3c) map to extended hydrogen-bond networks. These interface mutations support the structural observations that the well-packed hydrophobic core and precise surface structure of the interface are important for Rb function.

The Cys 706Phe mutation⁴, which is often used as an inactivating mutation in biochemical studies, occurs in the hydrophobic core of the B box, and would be expected to destabilize the folded state of the B box as there is no space for the larger Phe side chain.

Discussion

Phosphorylation sites of the pocket. The A and B boxes each contain a putative phosphorylation site with the minimal cyclin-dependent kinase consensus of Ser/Thr-Pro. On the A box, Ser 567 occurs near the A–B interface. The phosphorylation of Ser 567 appears unfavourable because its side chain is not solvent accessible, and its side-chain hydroxyl group makes two hydrogen bonds to backbone amide and carbonyl groups in a rigid portion of the structure (Fig. 3c).

Ser 780, whose phosphorylation *in vivo* has been suggested by recent data⁴⁰, is in the middle of a 13-residue carboxy-terminal tail of the B box (residues 775–787; Figs 1c, 2a). The tail, which is rich in proline and hydrophobic residues, packs against a hydrophobic patch on the cyclin fold of the B box. The Ser 780 side chain does not participate in any important interactions with the rest of the structure, but it is only partly solvent accessible. Ser 780 phosphorylation may thus involve changes in the tail structure that may increase the accessibility of the side chain.

Regulation of the LxCxE binding site by phosphorylation. The negative regulation of the LxCxE binding site is probably mediated by phosphorylation at the Thr 821 and Thr 826 sites⁴¹, about 30 residues after the B box. Two mechanisms have been proposed for this regulation⁴¹: that the phosphorylated polypeptide segment may bind to and block the site, or that phosphorylation causes a conformational change in the site. The structure does not favour the possibility of conformational changes because the LxCxE binding site has few flexible elements and it consists of conserved cyclin-fold

Table 2 Statistics from the crystallographic analysis

Data set	Native	Hg λ 1	Hg λ 2	Hg λ 3					
Wavelength (Å)	0.9080	1.0093	1.0060	0.9959					
Beam line	CHESS A1	NSLS-X4a	NSLS-X4a	NSLS-X4a					
Resolution (Å)	1.85	2.2	2.2	2.2					
Observations	135,820	101,930	101,113	102,808					
Unique reflections	45,533	24,328	24,288	24,368					
Data coverage (%)	91.8	95.1	95.2	95.7					
R_{sym} (%)	5.0	5.2	5.8	5.9					
MAD analysis (20.0–2.5 Å): phasing power	1.05	–	0.98	1.43					
R_{cullis}	0.89	–	0.91	0.75					
Anomalous R_{cullis}	–	0.73	0.70	0.81					
Mean FOM		0.64							
Refinement statistics:									
							r.m.s.d.		
Resolution (Å)	Reflections ($ F > 2\sigma$)	Protein atoms	Water atoms	R -factor (%)	R -free (%)	bonds (Å)	angles (°)	B-factor (Å ²)	
7.0–1.85	39,098	2,713	389	21.8	28.9	0.013	1.618	2.9	

$R_{\text{sym}} = \sum_i \sum_h |I_{hi} - \langle I_h \rangle| / \sum_i \sum_h I_{hi}$ for the intensity (I) of observations of reflection h . Phasing power = $(F_{\text{calc}})/E$ where (F_{calc}) is the root mean square heavy atom structure factor and E is the residual lack of closure error. R_{cullis} is the mean residual lack of closure error divided by the dispersive or anomalous difference. Mean figure of merit (FOM) = $\sum_i |F(hkl)|_{\text{obs}} / |F(hkl)|$. R -factor = $\sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R -free = R -factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement. r.m.s.d. root mean square deviations from ideal geometry and root mean square variation in the B -factor of bound atoms.

helices that have the same relative orientation in other structures. On the other hand, the structure supports the mechanism of direct competition as it reveals a six-lysine basic patch on the rim of the LxCxE binding site that may facilitate the binding of the phosphorylated polypeptide segment (Fig. 4d).

E2F binding. Several studies suggest that the E2F–Rb complex has several interacting regions^{10,16,42}. On E2F, the primary interaction

site appears to be the 18-residue Rb-binding motif, as it is necessary and sufficient for detectable complex formation¹⁴ and its mutation can abolish Rb binding²⁷. That a binding site for the 18-residue motif is contained within the Rb pocket is suggested by previous studies^{14,15} and by our isothermal titration calorimetry (ITC) binding data (Table 1). The structure of the Rb pocket suggests that the A–B interface is a possible E2F binding site because: (1) the A–B interface is the only portion of the pocket surface, apart from the LxCxE binding site, that is highly conserved (Figs 2a, 3d); (2) a subset of tumour-derived mutations map to interface residues, including to residues on the surface (Fig. 3a–c); and (3) there is a groove running half-way around the interface that could be suitable for peptide binding (Fig. 3d).

Studies have indicated that E2F and the LxCxE motif bind on separate sites on Rb^{43,44}. In support of this, the ITC binding data show that the affinity of the 18-residue E2F peptide for the Rb pocket is not affected by saturating amounts of the E7 LxCxE peptide (Table 1), and a ternary complex containing the Rb pocket, the E7 LxCxE peptide, and the 18-residue E2F peptide at roughly equimolar ratios can be isolated by gel filtration chromatography (data not shown). Dissociation of the Rb–E2F complex by E7 may thus involve competition between E2F and E7 for a site on Rb other than the LxCxE binding site⁴⁵.

Similarities with other cyclin-fold proteins. The cyclin fold of the B box is more similar to the cyclin folds in cyclinA³⁵ and TFIIB³⁶ (Fig. 5a; r.m.s. deviations of 1.6, 2.2, 1.4 and 1.7 Å from the cyclinA and TFIIB folds, respectively) than is the A-box cyclin fold (r.m.s. deviations of 2.1, 3.0, 1.9 and 2.3 Å). The lesser structural divergence of the B box, coupled to the B box having higher sequence conservation among Rb homologues, is consistent with it being the binding site for the LxCxE motif.

Although the biological functions and overall structures of Rb, cyclin A and TFIIB are different, they use their cyclin folds, and in particular the region wedged in between the third, fourth and fifth cyclin-fold helices, in similar ways in binding to their protein targets. Just as the B box uses this site to bind the LxCxE peptide, cyclinA uses it to bind the ‘PSTAIR’ helix of Cdk2 (ref. 35), and TFIIB uses it to bind a ‘stirrup’ of TBP³⁶ (Fig. 5b). It is noteworthy that this TBP stirrup provides a glutamic-acid residue that makes hydrogen-bond contacts³⁶ similar to those made by the glutamic acid of the LxCxE.

These similarities identify this region of the cyclin fold as particularly amenable to protein–protein interactions. In support of this, the corresponding regions in the second cyclin folds of cyclin A and TFIIB, which are not involved in intermolecular binding, are sequestered by extended polypeptide segments that bind in this site intramolecularly (Fig. 5b), and the corresponding region of the A box binds an A–B interface helix (Fig. 5b). □

a Cyclin folds from Rb(2), cyclin A (2), TFIIB (2)

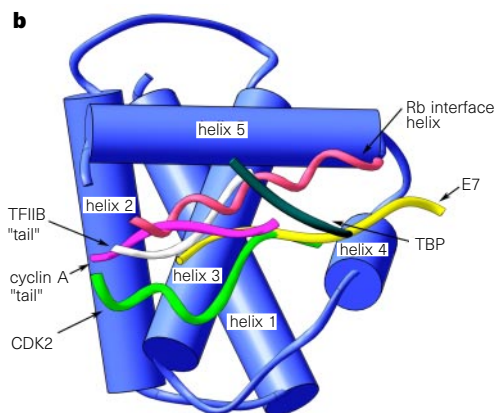
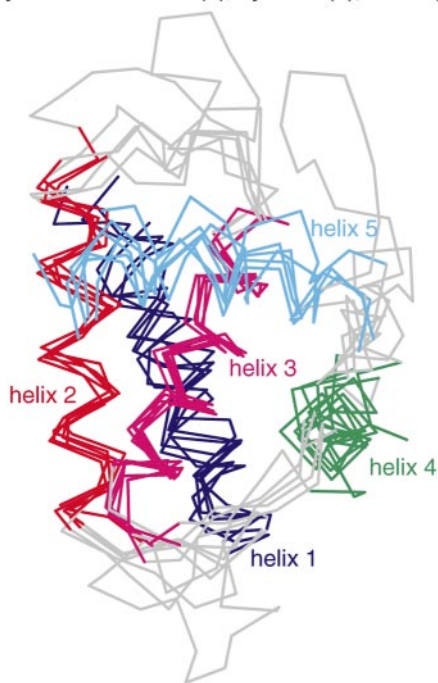


Figure 5 Rb, cyclin A and TFIIB use a common feature of their structurally conserved cyclin folds to bind their target proteins. **a**, Superposition of six cyclin folds, two each from Rb, cyclin A and TFIIB, shows that the relative arrangement of the five helices that make up the core of the cyclin fold is well conserved. The superposition was done by aligning 64 residues from the five core helices of each cyclin fold. The relative arrangement of the two repeats in each of the structures is different. **b**, Composite figure indicating the occupants of the LxCxE-binding site in the six superimposed cyclin folds from the Rb–E7, cyclin A–CDK2 and TFIIB–TBP complexes. This site is bound either intermolecularly by portions of target proteins: E7 for the Rb B box, Cdk2 for repeat 1 of cyclin A, and TBP for repeat 1 of TFIIB, or intramolecularly by structural elements from the same protein: an A–B interface helix from the Rb A box, or C-terminal ‘tails’ from the second repeats of cyclin A or TFIIB. For clarity, cyclin A’s repeat 1, which has few non-cyclin-fold elements, is shown.

Methods

Protein expression and purification. The recombinant human Rb pocket (residues 372 to 787) with two thrombin sites (LVPRGS) inserted after residues 589 and 635 was overexpressed at room temperature in *Escherichia coli* as a glutathione S-transferase (GST) fusion protein using the pGEX-2T vector (Pharmacia), and was purified by glutathione Sepharose and anion-exchange chromatography. Proteolytic digestion of a larger version of the Rb pocket (residues 364 to 928) was used in estimating the optimal boundaries of the pocket and the placement of the thrombin sites. Residues 590 to 635 of the spacer and the GST fusion protein were removed by digestion with thrombin (10 units per mg of Rb) at 4 °C without any notable protein loss. The resulting Rb pocket, the boundaries of which were confirmed by mass spectroscopy, was further purified by heparin Sepharose, glutathione Sepharose, and Superdex 200 gel filtration chromatography. The LxCxE peptide of HPV16 E7, which was synthesized chemically with the N and C termini acetylated and amidated, respectively, was purified by reversed phased chromatography, and it was mixed with the Rb pocket at a 2:1 molar ratio. The complex was purified by gel filtration and concentrated by ultrafiltration to 15 mg ml^{−1}. All other Rb

proteins were also expressed as GST fusion proteins and purified similarly.

Crystallization and data collection. Initial microcrystals, which could be obtained only in the absence of the spacer, were grown at 4 °C by the hanging-drop vapour diffusion method from 3.5 M sodium formate, 20 μ M CdCl₂, 0.1 M HEPES–Na, pH 7.5. Crystals suitable for diffraction were grown after multiple rounds of macroseeding; CdCl₂, which appeared to be essential for the spontaneous crystal growth, was not needed for the growth of seeded crystals. All diffraction data were collected using crystals flash frozen at –170 °C in 6.0 M sodium formate. The crystals belong to space group C2 with $a = 103.5$ Å, $b = 76.8$ Å, $c = 64.0$ Å, and $\beta = 90.9^\circ$. The native data set was collected at the A1 beam line of the Cornell High Energy Synchrotron Source (MacCHESS); the multiwavelength anomalous diffraction (MAD) data sets were collected at the X4A beam line of the National Synchrotron Light Source (NSLS), from a single crystal soaked in 5 mM thimerosal for 3 days and 1 mM ethyl mercury chloride for 1 day. Data were processed with the programs DENZO and SCALEPACK.

MAD analysis, model building and refinement. Experimental phases were obtained with MAD using the Hgλ1 data set as the native, the dispersive differences from the Hgλ2, Hgλ3 and the native data sets (with negative occupancies for the heavy-atom sites for the native data set), and the anomalous differences from the Hgλ1, Hgλ2 and Hgλ3 data sets. The phases, calculated with the program MLPHARE⁴⁶, had a mean figure of merit of 0.64 at 2.5 Å resolution. After density modification with DM⁴⁶, the electron density map was readily interpretable and the complete model was built at this stage with the program O⁴⁷. The model was refined with X-PLOR⁴⁸ with a 5% R-free set put aside before refinement. The final refined model contains residues 380–464, 472–499, 514–577, 645–785 of Rb, 21–29 of the E7 peptide, and 389 water molecules. The following regions do not have clear electron density and are presumed to be disordered: 372–379 and 786–787 from the N and C termini of the pocket, respectively, 578–589 and 636–644 flanking the spacer that was removed, 465–471 and 500–513 from two loops in between the α4 and α5 and in between the α6 and α7 A-box helices, respectively. The disordered regions overall have very low sequence conservation, except for residues 467 and 468 which are between the α4 and α5 helices.

Calorimetry. Binding constants of the E7 (residues 21 to 29) or E2F-1 (residues 409 to 426) peptides to the various Rb proteins were measured by isothermal titration calorimetry (ITC) using the Micro Calorimetry System (MicroCal). Briefly, 0.4 mM of each peptide was titrated into a 0.05 mM solution of the Rb protein, both in a buffer of 200 mM NaCl, 50 mM Tris-Cl, 5 mM DTT, pH 8.2, at 20 °C. Titration curves, typically of 20 points, were fitted to the two-state binding model using the program ORIGIN (MicroCal). Thermal denaturation curves were obtained by differential scanning calorimetry (DSC), with 0.06 mM solutions of Rb in the same buffer as that used in the ITC experiments. Thermal denaturation curves were fitted to the two-state transition model using the program ORIGIN (MicroCal), and gave $T_m = 42.9$ °C for the pocket with the spacer deleted, and $T_m = 46.9$ °C for the pocket with the native spacer. Both had $\Delta H/\Delta H_v \sim 1$ (ΔH_v calculated from fitting the data to a non-two-state model), although accurate thermodynamic parameters could not be obtained because of the irreversibility of the melting *in vitro*.

Received 1 October; accepted 1 December 1997.

- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Yandell, D. W. *et al.* Oncogenic point mutations in the human retinoblastoma gene: their application to genetic counseling. *New Engl. J. Med.* **321**, 1689–1695 (1989).
- Horowitz, J. M. *et al.* Point mutational inactivation of the retinoblastoma antioncogene. *Science* **243**, 937–940 (1989).
- Kaye, F. J., Kratzke, R. A., Ferster, J. L. & Horowitz, J. M. A single amino acid substitution results in a retinoblastoma protein defective in phosphorylation and oncoprotein binding. *Proc. Natl Acad. Sci. USA* **87**, 6922–6926 (1990).
- Onadin, Z., Hogg, A., Baird, P. N. & Cowell, J. K. Oncogenic point mutations in exon 20 of the RB1 gene in families showing incomplete penetrance and mild expression of the retinoblastoma phenotype. *Proc. Natl Acad. Sci. USA* **89**, 6177–6181 (1992).
- Dyson, N., Howley, P. M., Munger, K. & Harlow, E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* **243**, 934–937 (1989).
- zur Hausen, H. Papillomavirus infections: a major cause of human cancers. *Biochim. Biophys. Acta* **1288**, f55–f58 (1996).
- Bookstein, R., Shew, J. Y., Chen, P. L., Scully, P. & Lee, W. H. Suppression of tumorigenicity of human prostate carcinoma cells by replacing a mutated RB gene. *Science* **247**, 712–715 (1990).
- Huang, H. J. *et al.* Suppression of the neoplastic phenotype by replacement of the RB gene in human cancer cells. *Science* **242**, 1563–1566 (1988).
- Qin, X. Q., Chittenden, T., Livingston, D. M. & Kaelin, W. G. Jr Identification of a growth suppression domain within the retinoblastoma gene product. *Genes Dev.* **6**, 953–964 (1992).
- Goodrich, D. W., Wang, N. P., Qian, Y. W., Lee, E. Y. & Lee, W. H. The retinoblastoma gene product regulates progression through the G1 phase of the cell cycle. *Cell* **67**, 293–302 (1991).

- Hinds, P. W. *et al.* Regulation of the retinoblastoma protein functions by ectopic expression of human cyclins. *Cell* **70**, 993–1006 (1992).
- Buchkovich, K., Duffy, L. A. & Harlow, E. The retinoblastoma protein is phosphorylated during specific phases of the cell cycle. *Cell* **58**, 1097–1105 (1989).
- Helin, K. *et al.* A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell* **70**, 33–50 (1992).
- Kaelin, W. G. Jr *et al.* Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell* **70**, 351–364 (1992).
- Hiebert, S. W., Chellappan, S. P., Horowitz, J. M. & Nevins, J. R. The interaction of RB with E2F coincides with an inhibition of the transcriptional activity of E2F. *Genes. Dev.* **6**, 177–185 (1992).
- Flemington, E. K., Speck, S. H. & Kaelin, W. G. Jr E2F-1-mediated transactivation is inhibited by complex formation with the retinoblastoma susceptibility gene product. *Proc. Natl Acad. Sci. USA* **90**, 6914–6918 (1993).
- Helin, K., Harlow, E. & Fattaey, A. Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. *Mol. Cell. Biol.* **13**, 6501–6508 (1993).
- Weintraub, S. J. *et al.* Mechanism of active transcriptional repression by the retinoblastoma protein. *Nature* **375**, 812–815 (1995).
- Sellers, W. R., Rodgers, J. W. & Kaelin, W. G. Jr A potent transrepression domain in the retinoblastoma protein induces a cell cycle arrest when bound to E2F sites. *Proc. Natl Acad. Sci. USA* **92**, 11544–11548 (1995).
- Herber, R., Liem, A., Pitot, H. & Lambert, P. F. Squamous epithelial hyperplasia and carcinoma in mice transgenic for the human papillomavirus type 16 E7 oncogene. *J. Virol.* **70**, 1873–1881 (1996).
- Chellappan, S. *et al.* Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proc. Natl Acad. Sci. USA* **89**, 4549–4553 (1992).
- Munger, K. *et al.* Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J.* **8**, 4099–4105 (1989).
- DeCaprio, J. A. *et al.* SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. *Cell* **54**, 275–283 (1988).
- Whyte, P. *et al.* Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature* **334**, 124–129 (1988).
- Gulliver, G. A., Herber, R. L., Liem, A. & Lambert, P. F. Both conserved region 1 (CR1) and CR2 of the human papillomavirus type 16 E7 oncogene are required for induction of epidermal hyperplasia and tumor formation in transgenic mice. *J. Virol.* **71**, 5905–5914 (1977).
- Shan, B., Durfee, T. & Lee, W. H. Disruption of RB/E2F-1 interaction by single point mutations in E2F-1 enhances S-phase entry and apoptosis. *Proc. Natl Acad. Sci. USA* **93**, 679–684 (1996).
- Hu, Q. J., Dyson, N. & Harlow, E. The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. *EMBO J.* **9**, 1147–1155 (1990).
- Kaelin, W. G. Jr, Ewen, M. E. & Livingston, D. M. Definition of the minimal simian virus 40 large T antigen- and adenovirus E1A-binding domain in the retinoblastoma gene product. *Mol. Cell. Biol.* **10**, 3761–3769 (1990).
- Huang, S., Wang, N. P., Tseng, B. Y., Lee, W. H. & Lee, E. H. Two distinct and frequently mutated regions of retinoblastoma protein are required for binding to SV40 T antigen. *EMBO J.* **9**, 1815–1822 (1990).
- Jones, R. E. *et al.* Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *J. Biol. Chem.* **265**, 12782–12785 (1990).
- Chow, K. N. & Dean, D. C. Domains A and B in the Rb pocket interact to form a transcriptional repressor motif. *Mol. Cell. Biol.* **16**, 4862–4868 (1996).
- Kim, H. Y. & Cho, Y. Structural similarity between the pocket region of retinoblastoma tumour suppressor and the cyclin-box. *Nature Struct. Biol.* **4**, 390–395 (1997).
- Gibson, T. J., Thompson, J. D., Blocker, A. & Kouzarides, T. Evidence for a protein domain superfamily shared by the cyclins, TFIIB and RB/p107. *Nucleic Acids Res.* **22**, 946–952 (1994).
- Jeffrey, P. D. *et al.* Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313–320 (1995).
- Nikolov, D. B. *et al.* Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
- Graf, G. *et al.* A maize cDNA encoding a member of the retinoblastoma protein family: involvement in endoreduplication. *Proc. Natl Acad. Sci. USA* **93**, 8962–8967 (1996).
- Dyson, N., Buchkovich, K., Whyte, P. & Harlow, E. The cellular 107K protein that binds to adenovirus E1A also associates with the large T antigens of SV40 and JC virus. *Cell* **58**, 249–255 (1989).
- Kratzke, R. A. *et al.* Partial inactivation of the RB product in a family with incomplete penetrance of familial retinoblastoma and benign retinal tumors. *Oncogene* **9**, 1321–1326 (1994).
- Kitagawa, M. *et al.* The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2. *EMBO J.* **15**, 7060–7069 (1996).
- Knudsen, E. S. & Wang, J. Y. J. Differential regulation of retinoblastoma protein function by specific Cdk phosphorylation sites. *J. Biol. Chem.* **271**, 8313–8320 (1996).
- O'Connor, R. J. & Hearing, P. Mutually exclusive interaction of the adenovirus E4-6/7 protein and the retinoblastoma gene product with internal domains of E2F-1 and DP-1. *J. Virol.* **68**, 6848–6862 (1994).
- Wu, E. W., Clemens, K. E., Heck, D. V. & Munger, K. The human papillomavirus E7 oncoprotein and the cellular transcription factor E2F bind to separate sites on the retinoblastoma tumor suppressor protein. *J. Virol.* **67**, 2402–2407 (1993).
- Ikedo, M. A. & Nevins, J. R. Identification of distinct roles for separate E1A domains in disruption of E2F complexes. *Mol. Cell. Biol.* **13**, 7029–7035 (1993).
- Fattaey, A. R., Harlow, E. & Helin, K. Independent regions of adenovirus E1A are required for binding to and dissociation of E2F-protein complexes. *Mol. Cell. Biol.* **13**, 7267–7277 (1993).
- Collaborative Computational Project B. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zhou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *X-PLOR, a System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1991).

Acknowledgements. We thank S. Geromanos and H. Erdjument-Bromage of the Sloan-Kettering Microchemistry Facility for N-terminal sequence and mass spectroscopic analyses; M. Ewen, T. D. Gilmore, W. Harper, M. H. Lee and J. Y. J. Wang for cDNA clones; and C. Ogata of the National Synchrotron Light Source X4A beam line and the staff of the Cornell High Energy Synchrotron Source MacChess for help with data collection. This work was supported by the NIH, the Howard Hughes Medical Institute, the Pew Charitable Trusts, the Arnold and Mabel Beckman Foundation, the Dewitt Wallace Foundation and the Samuel and May Rudin Foundation.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: Nikola@xray2.mskcc.org). Coordinates have been deposited with the Brookhaven Protein Data Bank (accession number 1GUX).

A population of faint galaxies that contribute to the cosmic X-ray background

Y. Ueda*, T. Takahashi*, H. Inoue*, T. Tsuru†, M. Sakano†, Y. Ishisaki‡, Y. Ogasaka§, K. Makishima||, T. Yamada¶, K. Ohta# & M. Akiyama#

* Institute of Space and Astronautical Science, Yoshinodai, Sagami-hara, Kanagawa 229, Japan

† Department of Physics, Kyoto University, Kyoto 606-01, Japan

‡ Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-03, Japan

§ NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

|| Department of Physics, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

¶ Astronomical Institute, Tohoku University, Sendai 980-77, Japan

Department of Astronomy, Kyoto University, Kyoto 606-01, Japan

The origin of the cosmic X-ray background radiation^{1,2} has remained mysterious since its discovery³ thirty-five years ago. Investigation of its origin has been difficult because instruments have had insufficient resolution to distinguish small, faint sources in the hard X-ray band (above 2 keV) that dominates the background. Until now, only three per cent of the flux in the 2–10 keV band could be attributed to individual sources^{4,5}. Here we report the results of a survey 100 times more sensitive than previous studies in the 2–10 keV band. We find many faint resolved sources, whose integrated flux accounts for 30 per cent of the X-ray background in this energy range. The average spectrum of the resolved sources is harder than those of nearby bright active galactic nuclei and is close to the spectrum of the X-ray background radiation. This means that a new class of sources, with hard X-ray spectra, dominate the sky at photon energies above 2 keV.

The origin of the cosmic X-ray background (CXB) has been a long-standing puzzle in X-ray astronomy. Its spectrum resembles that of a hot plasma with a temperature of 40 keV (ref. 6), which introduced the interesting possibility that hot plasmas uniformly fill the universe. However, little distortion of the cosmic microwave background (CMB) spectrum detected by the COBE experiment⁷ contradicts this possibility and indicates that the CXB is most likely to be a superposition of faint discrete sources. In the soft X-ray band, the Rosat satellite resolved 60% of the CXB in the 0.5–2 keV band into individual sources: the majority are active galactic nuclei (AGNs) such as quasars^{2,8}. By contrast, in the hard X-ray band above 2 keV, resolved sources only accounted for 3% of the total emission. In addition, there has been a major problem called the ‘spectral paradox’: AGNs observed with HEAO1, EXOSAT and Ginga have on average a power-law photon index Γ of 1.7–1.9 in the 2–10 keV band^{9,10}, softer than that of the CXB, $\Gamma \sim 1.4$ (refs 6, 11). This indicates that faint counterparts of these AGNs cannot be the principal contributors to the CXB, and that other classes of source with harder spectra should exist at a fainter flux level.

To study the nature of the hard X-ray source population and thus provide a clear picture of the composition of the CXB, the first unbiased imaging sky survey (Large Sky Survey, LSS) above 2 keV (up to 10 keV) has been performed with the ASCA satellite¹². ASCA carries four identical X-ray telescopes (XRT)¹³, which are coupled to two solid-state imaging spectrometers (SIS) and two gas imaging spectrometers (GIS)¹⁴ as focal plane detectors. The overall ASCA sensitivity spans a wide energy range from 0.5 to 10 keV. The sensitivity of ASCA, higher than any previous missions by two orders of magnitude in the 2–10 keV band, enables us to investigate the statistical properties of X-ray sources at a flux level to $\sim 10^{-13}$ erg s^{-1} cm^{-2} (2–10 keV).

The survey field is a continuous region, centred at RA = 13 h 14 min, dec. = 31° 30' (equinox 2000), near the north galactic pole. Seventy-six pointings were made onto the LSS region over several periods from December 1993 to July 1995. The pointing positions were arranged so that the sum of field of views (FOVs) of the SIS (22' $\times$ 22') covers continuously with good overlaps. Total

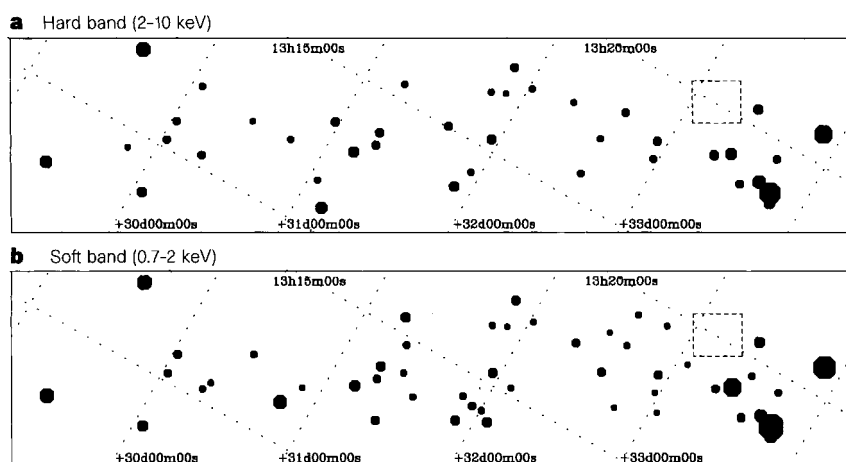


Figure 1 ASCA Large Sky Survey (LSS). Sources detected in the LSS field with the GIS in 2–10 keV (top) and in 0.7–2 keV (bottom). The radius of each circle is proportional to the flux, ranging from: top 1.1×10^{-13} to 1.4×10^{-12} erg s^{-1} cm^{-2} (2–10 keV); and bottom, 5.2×10^{-14} to 1.1×10^{-12} erg s^{-1} cm^{-2} (0.5–2 keV). Ratio of radius between the two panels indicates the hardness of the sources such that a source with a photon index of 1.7 has the same radius. A photon index of 1.7 and N_H of 1.1×10^{20} cm^{-2} is assumed to convert the count rate to the flux in each energy band. The square surrounded by dashed lines represents a region where our analysis was not performed to avoid systematic errors due to a bright extended source located there. We carried out the source-detection procedure in two steps. In the first step, raw images are cross-correlated (smoothed) with the position-dependent PSF, and source candidates are searched in the

smoothed image by finding local peaks. In the second step, we performed a two-dimensional fitting to the raw image in the photon count space using the Poisson maximum likelihood algorithm (T.T. *et al.*, manuscript in preparation). The model function consists of the background (CXB + NXB) and source peaks found in the first step. The significance of each detected source is defined as $\sigma = (\text{best-fit flux}) / (1\sigma \text{ statistical error of the flux})$. We adopted the following criteria for source detection: (1) the GIS significance is above 3.5σ , and (2) the significance for the summed count rate of the GIS and the SIS exceeds 4.5σ . With these criteria, we found that the contamination by fake sources is less than a few per cent of those detected overall. Uncertainties in the absolute flux are of the order of 10% (ref. 22).

sky area observed with the GIS and SIS amounts to 7.0 deg^2 and 5.4 deg^2 . The mean exposure time per point is 54 ks (GIS) and 27 ks (SIS). Details of data selection criteria are described in ref. 15.

Despite the high sensitivity of ASCA, careful analysis is required to study faint sources that have small numbers of photons in the detector plane. Particularly, the point spread function (PSF) of the XRT has a broad wing¹³, which affects flux determination of a faint source near a brighter one. Moreover, the PSF strongly depends on the position in the focal plane: the larger the off-axis angle is, the more distorted the PSF becomes. The background (the unresolved CXB and the non-X-ray background (NXB)) also depends on the position and energy. To overcome these difficulties and make the best use of the ASCA capability, we have developed a new source-detection method¹⁵, in which we took account of the complicated responses of the XRT and the detectors. The source-finding procedure consists of two steps: source detection and flux calculation. In each step, we analysed the GIS and the SIS data separately and then combined them, taking advantage of each detector: the GIS has a superior detection efficiency above 2 keV and a wider FOV, whereas the SIS with the XRT performs well in separating neighbouring sources by the angular resolution of 1 arcmin.

We applied the method to the combined image of the LSS field, independently in the hard (2–10 keV) and the soft (0.7–2 keV) band. Figure 1 shows an X-ray source map of the LSS field in the two energy bands. We detected 40 sources in the hard band and 50 sources in the soft band. Among them, 31 sources are detected in both bands. The 2–10 keV fluxes of the hard band detected sources distributed in a wide range from $1.1 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ to $1.4 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV). It is noticeable that nine sources were detected only in the hard band.

The contribution of the sources to the CXB can be calculated from the cumulative source number count, $N(>S)$, as a function of flux S . Figure 2 shows the $\log N - \log S$ relation obtained from the LSS data in the 2–10 keV band. Our result lies on the extrapolation from the previous results with $N(>S) \propto S^{-3/2}$, which is expected

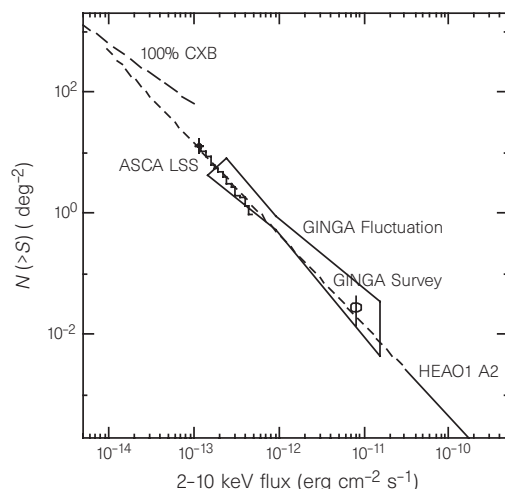


Figure 2 The $\log N - \log S$ relation in the hard band (2–10 keV). The steps around $(1-5) \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ represent the present results obtained from the 2–10 keV survey: we confirmed that associated systematic errors, such as the effects of source confusion, were less than 20% by using Monte Carlo simulations. A photon index of 1.5 is assumed when converting count rate to flux. The thick solid line above $3 \times 10^{-11} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV) represents the $\log N - \log S$ relation obtained from the HEAO1 A2 (ref. 4). The 'bow tie' shape represents constraints from the fluctuation analysis by Ginga²³. Source counts by Ginga survey⁵ are also plotted as a circle. The short-dashed line represents extrapolation with a slope of -1.5 from the HEAO1 A2 result. The long-dashed line corresponds to the condition that the total intensity of the CXB is explained by the integral of sources, assuming $N(>S)$ is proportional to $S^{-1.5}$.

from a uniform distribution of sources in euclidean space. The integrated intensity calculated from the $\log N - \log S$ relation above a flux of $\sim 1 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV) accounts for 25–30% of the CXB intensity in the 2–10 keV band^{6,11}. Note that the $\log N - \log S$ relation from the 0.7–2 keV band survey is consistent with the previous results from Einstein¹⁶ and Rosat¹⁷, where our sensitivity limit is $\sim 5 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ in this energy band.

To confirm that our result is not much affected by the source confusion, we performed a Monte Carlo simulation. We developed an instrument simulator that fully simulates the complicated responses of ASCA. As the input for the simulator, the sky image was created according to the $\log N - \log S$ relation obtained by the present study. We then analysed simulated data and applied our source-finding procedure in exactly the same manner as for the real data. For sources detected, we compared the observed flux with the input flux, which we defined as a flux of the brightest source within a typical error radius ($1'$) around the detected position. We found a weak tendency that observed fluxes become larger than input ones when the flux is below $\sim 1 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV), which is attributed to the source confusion by neighbouring sources. However, their contamination of the observed flux of resolved sources was estimated to be less than 20%. We conclude that the effect of the source confusion does not greatly affect our results. We verified that the $\log N - \log S$ relation derived from the simulated data is fully consistent with the input parameters.

The most important point related to the origin of the CXB is the spectrum of X-ray sources at much fainter flux limit than achieved by previous observations. For the detected sources, we calculated fluxes in 2–4 keV and 4–10 keV to determine their spectra. We summed up the spectra of 32 sources with fluxes from 1×10^{-13} to $4 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV) to obtain the average spectrum. Assuming a power law absorbed with a galactic neutral hydrogen column density of $N_{\text{H}} = 1.1 \times 10^{20} \text{ cm}^{-2}$, we obtained a photon index of 1.52 ± 0.17 (90% statistical error in the mean value). Figure 3 shows the relation between the flux and the average photon

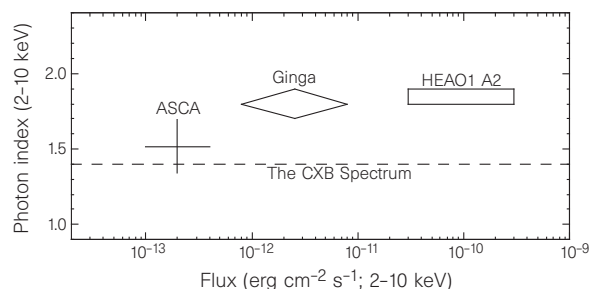


Figure 3 Relation between the flux and the average photon index in the 2–10 keV band. The cross represents our result derived from the sources detected in the 2–10 keV band with fluxes from 1×10^{-13} to $4 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV). The flux of each source is converted from the count rate, assuming a photon index of 1.5. The diamond represents the result from the Ginga fluctuation analysis²³, one decade below the sensitivity limit of the Ginga survey ($8 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$ in 2–10 keV). The rectangular box shows the result of the Piccinotti sample by HEAO1 A-2 (ref. 4). It consists of 30 AGNs and 30 clusters of galaxies. The typical spectrum of these clusters roughly corresponds to a photon index of ~ 2 if we assume a thermal emission with a temperature of 6 keV.

index in 2–10 keV. The dashed line corresponds to the photon index of the CXB spectrum in the same energy band, 1.4 (refs 6, 11). The average photon index of the faint sources detected in our survey is significantly harder than that obtained from previous measurements with a much higher flux level in this energy band. Note that eight sources brighter than $4 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2–10 keV) show an average photon index of 2.1 ± 0.4 , which is consistent with the previous results.

This suggests that a solution to the spectral paradox, the biggest puzzle in X-ray astronomy, is emerging. Our results indicate that a population of sources with an average photon index similar to the CXB begins to dominate at energies above 2 keV at the sensitivity we achieved.

What are the sources detected in our survey? The average spectrum obtained from the hard-band sample is harder than that of nearby type-I AGNs found at a flux level of $\sim 10^{-11} \text{ erg cm}^{-2} \text{ cm}^{-2}$ (2–10 keV). Combined with previous studies made in the soft X-ray band, we can estimate the fraction of type-I AGNs in the detected sources. If we assume that these AGNs have $\Gamma = 2.0$ below 2 keV (ref. 17) and Γ values of 1.7–1.9 above 2 keV (refs 9, 10) as a typical spectrum of type-I AGNs, our sensitivity limit of $10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ in 2–10 keV corresponds to $(7\text{--}8) \times 10^{-14} \text{ erg s}^{-2}$ in the 0.5–2 keV band. According to the Rosat deep surveys² and Einstein EMSS¹⁸, the cumulative source counts is 5–6 deg⁻² and ~60% of these sources are identified as type-I AGNs at this flux level. Thus, the fraction of type-I AGNs in our sample is calculated to be only 20–30%. To have an average spectrum with $\Gamma = 1.52 \pm 0.17$ above 2 keV, the remaining 70–80% of sources should have an average index of $\Gamma = 1.45 \pm 0.25$.

With our sensitivity limit, the number of these hard X-ray sources already exceeds the population of type-I AGNs. The highly absorbed AGNs, so-called type-II AGNs, are good candidates for this population of objects. The presence of a large number of type-II AGNs is expected from the so-called ‘unified scheme’ of AGNs^{19,20}, where the direction of line of sight determines the apparent properties of AGNs. If we directly see the emission from the nucleus, it appears as a type-I AGN; if the matter around the nucleus blocks our line of sight (with a column density of $N_{\text{H}} \sim 10^{22-24} \text{ cm}^{-2}$), we will see a highly absorbed spectrum, which is hard in appearance and considered as a type-II AGN. The deep X-ray follow-up observations of the hardest source in the LSS field, AXJ 131501 + 3141, revealed a highly absorbed X-ray spectrum, with N_{H} of $6 \times 10^{22} \text{ cm}^{-2}$ (ref. 24). It has been optically identified as a type-II Seyfert galaxy at $z = 0.072$, with an intrinsic X-ray luminosity of $2 \times 10^{43} \text{ erg s}^{-1}$ (2–10 keV) (ref. 21). It would be interesting to investigate the possible connection of these sources with the soft X-ray population identified as narrow emission line galaxies (NELGs) in the Rosat surveys², although the overall properties of NELGs have not yet been established. The results presented here represent an important step towards fully understanding the origin of the CXB. Future X-ray missions in the hard X-ray band will be important for reaching a final solution of the CXB. □

Received 3 April; accepted 9 December 1997.

1. Fabian, A. C. & Barcons, X. The origin of the X-ray background. *Annu. Rev. Astron. Astrophys.* **30**, 429–456 (1992).
2. Hasinger, G. The extragalactic X-ray and gamma-ray background. *Astron. Astrophys. Suppl. Ser.* **120**, 607–614 (1996).
3. Giacconi, R., Grusky, H., Paolini, F. R. & Rossi, B. B. Evidence for X-rays from sources outside the Solar system. *Phys. Rev. Lett.* **9**, 439–443 (1962).
4. Piccinotti, G. *et al.* A complex X-ray sample of the high-latitude ($|b| > 20^\circ$) sky from HEAO 1 A-2: $\text{Log} N - \text{Log} S$ and luminosity functions. *Astrophys. J.* **253**, 485–503 (1982).
5. Kondo, H. *et al.* in *Proc. Frontiers of X-ray Astronomy* (eds Tanaka, Y. & Koyama, K.) 655–656 (Universal Academy, Tokyo, 1991).
6. Marshall, F. *et al.* The diffuse X-ray background spectrum from 3 to 50 keV. *Astrophys. J.* **235**, 4–10 (1980).
7. Mather, J. C. *et al.* Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *Astrophys. J.* **420**, 439–444 (1994).
8. Shanks, T. *et al.* The origin of the cosmic X-ray background. *Nature* **353**, 315–320 (1991).
9. Turner, T. J. & Pounds, K. A. The EXOSAT spectral survey of AGN. *Mon. Not. R. Astron. Soc.* **240**, 833–880 (1989).
10. Williams, O. R. *et al.* The X-ray spectra of high-luminosity active galactic nuclei observed by *Ginga*.

- Astrophys. J.* **389**, 157–178 (1992).
11. Gendreau, K. C. *et al.* ASCA observations of the spectrum of the X-ray background. *Publ. Astron. Soc. Jpn* **47**, L5–L10 (1995).
12. Tanaka, Y., Inoue, H. & Holt, S. S. The X-ray astronomy satellite ASCA. *Publ. Astron. Soc. Jpn* **46**, L37–L41 (1994).
13. Serlemitsos, P. J. *et al.* The X-ray telescope on board ASCA. *Publ. Astron. Soc. Jpn* **47**, 105–114 (1995).
14. Ohashi, T. *et al.* The gas imaging spectrometer on board ASCA. *Publ. Astron. Soc. Jpn* **48**, 157–170 (1996).
15. Ueda, Y. *ASCA Studies of Faint X-ray Sources and the Relation to the Cosmic X-ray Background*. Thesis, Univ. Tokyo (1996).
16. Giacconi, R. H. *et al.* A high-sensitivity X-ray survey using the Einstein Observatory and the discrete source contribution to the extragalactic X-ray background. *Astrophys. J.* **234**, L1–L7 (1979).
17. Hasinger, G. *et al.* A deep X-ray survey in the Lockman Hole and the soft X-ray log $N - \log S$. *Astron. Astrophys.* **275**, 1–15 (1993).
18. Stocke, J. T. *et al.* The Einstein Observatory extended medium-sensitivity survey. II—the optical identifications. *Astrophys. J. Suppl. Ser.* **76**, 813–874 (1991).
19. Antonucci, R. R. J. & Miller, J. S. Spectropolarimetry and the nature of NGC 1068. *Astrophys. J.* **297**, 621–632 (1985).
20. Awaki, H., Koyama, K., Inoue, H. & Halpern, J. P. X-ray implications of a unified model of Seyfert galaxies. *Publ. Astron. Soc. Jpn* **43**, 195–212 (1991).
21. Akiyama, M. *et al.* Optical identification of the hardest X-ray source in the ASCA Large Sky Survey. *Astrophys. J.* (in the press).
22. Makishima, K. *et al.* In-orbit performance of the gas imaging spectrometer onboard ASCA. *Publ. Astron. Soc. Jpn* **48**, 171–190 (1996).
23. Hayashida, K., Inoue, H. & Kii, T. in *Proc. Frontiers of X-ray Astronomy* (eds Tanaka, Y. & Koyama, K.) 653–654 (Universal Academy, Tokyo, 1991).
24. Sakano, M. *et al.* The hardest X-ray source in the ASCA Large Sky Survey: Discovery of a new type 2 Seyfert. *Astrophys. J.* (submitted).

Acknowledgements. We acknowledge the ASCA_ANL and SimASCA software development teams for supporting the analysis technique. We thank the members of the ASCA team for spacecraft operation and data acquisition, and T. Kii and members of the LSS team for discussion.

An oxygen-rich dust disk surrounding an evolved star in the Red Rectangle

L. B. F. M. Waters*†, C. Waelkens‡, H. van Winckel‡, F. J. Molster*, A. G. M. Tielens‡§, J. Th. van Loon*, P. W. Morris||, J. Cami*†, J. Bouwman*, A. de Koter*, T. de Jong*† & Th. de Graauw†

* Astronomical Institute ‘Anton Pannekoek’, University of Amsterdam, Kruislaan 403, NL-1098 SJ Amsterdam, The Netherlands

† SRON Groningen, and § Kapteyn Astronomical Institute, PO Box 800, NL-9700 AV Groningen, The Netherlands

‡ Instituut voor Sterrenkunde, K.U. Leuven, Celestijnenlaan 200B, 3001 Heverlee, Belgium

|| SRON Utrecht, Sorbonnelaan 2, 3582 CA Utrecht, The Netherlands

The Red Rectangle¹ is the prototype of a class of carbon-rich reflection nebulae surrounding low-mass stars in the final stages of evolution. The central star of this nebula has ejected most of its layers (during the red-giant phase), which now form the surrounding cloud, and is rapidly evolving to a white dwarf. This star is also a member of a wide binary system², which is surrounded by a thick, dusty disk of material^{3,4}. Here we report infrared observations of the Red Rectangle that reveal the presence of oxygen-rich material: prominent emission bands from crystalline silicates, and absorption lines arising from carbon dioxide. The oxygen-rich material is located in the circumbinary disk, in contrast to the previously known carbon-rich dust, which is found mainly in the extended nebula^{5,6}. The properties of the oxygen-rich dust are similar to those of dusty disks surrounding young stars⁷, which are believed to be the sites of planet formation. Grain processing, and perhaps even planet formation, may therefore also be occurring in the circumbinary disk of this evolved star.

The Red Rectangle nebula¹ was discovered by Cohen and co-workers in 1975 as a peculiar X-shaped nebulosity. High-resolution optical and near-infrared imaging reveals the presence of an optically thick disk in the centre of the X-shaped nebula, with scattering lobes extending above and below the disk^{3,4}. The infrared spectrum of the nebula is dominated by emission from the well

known unidentified infrared (UIR) bands⁸ at 3.29, 6.2, 7.7, 8.6 and 11.3 μm , usually attributed to polycyclic aromatic hydrocarbons⁹ (PAHs), and proving the rich carbon-based chemistry of the nebula. Infrared imaging of the UIR bands shows that the carriers of these bands are located in the nebula and not in the central part (disk)^{5,6}. The primary of the binary system HD44179 has an effective temperature of 7,500 K, and has a photosphere which is depleted of refractory elements but has solar C, N, O, S and Zn abundances¹⁰. This depletion pattern is similar to that of the gas component of the interstellar medium (ISM) and is the result of fractionation onto dust, and subsequent re-accretion of gas devoid of metals onto the star^{11,12}.

The origin of the nebula is related to mass loss and/or mass transfer in a low-mass close binary system, of which the more massive component evolved to the asymptotic giant branch (AGB). With an orbital period of 318 days (ref. 10), the system is too small to accommodate an AGB star, and mass transfer must have taken place, rapidly circularizing the orbit^{10,4}. The present-day eccentricity of 0.38 probably resulted from tidal interaction with the massive circumbinary disk, similar to that seen in young binary systems¹³.

We observed the Red Rectangle and its central binary on 18 October, 1997 with the short wavelength spectrometer¹⁴ (SWS) of the Infrared Space Observatory¹⁵ (ISO). The spectrum covers the entire SWS wavelength range from 2.38 to 45.2 μm . We show the full spectrum in Fig. 1, and the 15 μm and 4.5 μm (continuum subtracted) region is displayed in Fig. 2. The λ 2.4–15 μm spectrum is characterized by a rising continuum with superimposed very strong emission from the UIR bands, originating from the X-shaped nebula^{5,6} (Fig. 3). We report here the discovery of two new bands in the 13–15 μm region of the Red Rectangle spectrum (Fig. 1) at λ 13.57 and 14.23 μm . This region contains the spectral signatures of PAH-edge structures. The new bands may be identified with PAHs

containing four (λ 13.57 μm) and five (λ 14.23 μm) adjacent H atoms. A band near 13.58 μm is also seen in NGC7027, and near 13.48 μm in IRAS21282 + 5050¹⁶. These bands may be related to the band at 13.57 μm seen in the Red Rectangle. A more detailed description of the PAH spectrum will be presented elsewhere. Here we concentrate on the presence of oxygen-rich dust and gas components.

The spectrum at wavelengths larger than 15 μm is dominated by numerous narrow emission bands which can be identified with crystalline silicates^{7,17}. Peaks from olivines¹⁸ ($(\text{Mg}_x\text{Fe}_{1-x})_2\text{SiO}_4$) are found at 19.56, 23.75, 27.78 and 33.85 μm , and from pyroxenes¹⁹ ($\text{Mg}_x\text{Fe}_{1-x}\text{SiO}_3$) at 29.5, 33.0, 36.22, 40.56 and possibly at 43.99 μm . The last feature may also contain a contribution from crystalline H_2O ice²⁰. The spectra show a remarkable degree of fine-structure in the bands (see Fig. 1), not seen previously, suggesting that several carriers contribute to the bands. The band strength ratio of the 23.75 and 33.85 μm can be used to derive the temperature of the carriers of these bands, and we find it to be ~ 120 K. We have fitted a simple dust model to the spectrum and find approximate values for the inner and outer radii of the oxygen-rich dust component of 500 and 2,000 AU, respectively (using a distance of 330 pc, a stellar radius of $40 R_\odot$ and stellar temperature of 7,500 K), and a dust mass of $4 \times 10^{-5} M_\odot$, of which about 13 per cent is crystalline. The dust mass of the carbon-rich component is a factor of 10 smaller. The model severely underestimates the millimetre continuum measurements²¹, and another component, with large grains, must be present. The millimetre data yield a dust mass of $6 \times 10^{-4} M_\odot$ (ref. 21). This is probably a lower limit because of the presence of very large grains²². We find a radial density gradient for the oxygen-rich dust component of $n(r) \propto r^{-1.0}$, but this value is poorly constrained. Nevertheless it suggests that the material is not in an outflow.

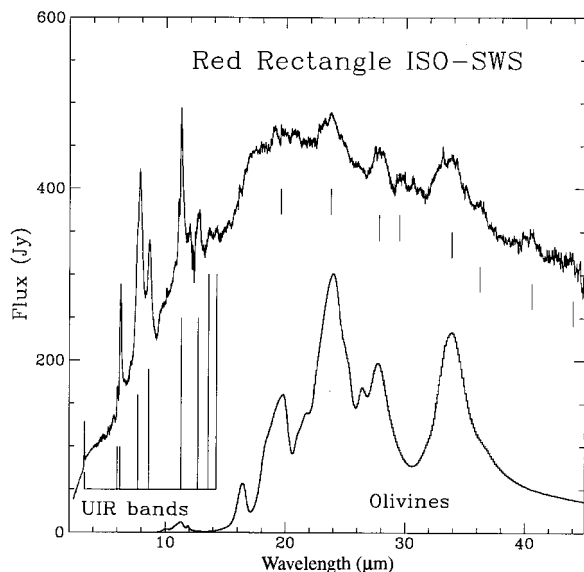


Figure 1 Full SWS grating scan (AOT01 speed 4) of the Red Rectangle nebula and its central star. The SWS apertures (14×20 , 14×27 and 20×33 arcsec) cover most of the nebulosity. Data reduction was standard and done with version 6 of the interactive analysis system. The data of all 12 detectors per AOT sub-band were sigma-clipped, averaged and rebinned to a uniform resolution of $\lambda/\Delta\lambda = 1,000$, which is close to the spectral resolution of this observing mode. The AOT sub-bands were then joined to form a single spectrum covering λ 2.38 to 45.2 μm . The thick tickmarks indicate the positions of the strongest oxygen-rich bands. The lower solid line gives the emissivity of olivines with $T = 120$ K. Note also the presence of strong emission from the unidentified infrared bands (UIRs) at $\lambda < 15 \mu\text{m}$ (indicated by the thin tickmarks), usually attributed to polycyclic aromatic hydrocarbons (PAHs). The presence of both carbon-rich and oxygen-rich chemistry is related to the mass-loss history of the central binary system.

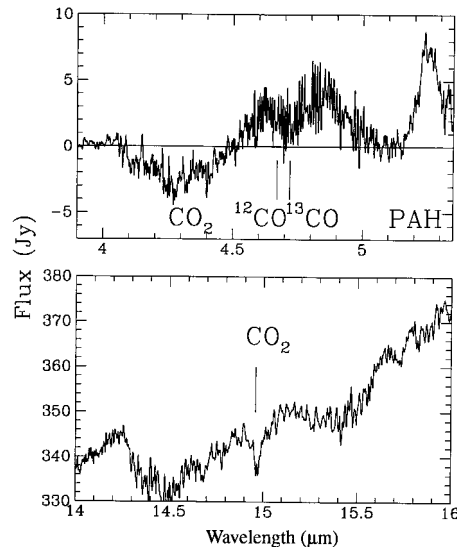


Figure 2 Expanded view of the spectral region around CO_2 absorption features. Top, SWS spectrum in the 4.5 μm region, showing weak absorption from CO_2 , emission from gas-phase ^{12}CO and ^{13}CO , and the 5.3 μm PAH feature. Bottom, SWS spectrum in the 15 μm region with weak 15 μm gas-phase CO_2 absorption.

At 4.27 and 15 μm , we find weak absorption from gas-phase CO_2 (Fig. 2). The 4.27 μm band may also have a contribution from solid CO_2 . Given the large circumstellar extinction towards the central star, the CO_2 absorption must be of circumstellar origin. We also find emission from the gas-phase ^{12}CO and ^{13}CO fundamental rotational/vibrational bands (Fig. 2), for which an excitation temperature of about 300–400 K is found. The excitation temperature and emission character suggest that the CO responsible for the emission at 4.6 μm is located in the nebula.

The presence of oxygen-rich material in the Red Rectangle, already suggested on the basis of OH absorption at ultraviolet wavelengths²³, is now convincingly demonstrated by our detection of CO_2 absorption and crystalline silicate emission. The OH and CO_2 absorption show that the oxygen-rich material is in the absorbing material towards the central star. It is therefore reasonable to assume that the oxygen-rich material is located in the circum-binary disk. The lack of the carbon-rich UIR emission in the central part of the nebula^{5,6} (see Fig. 3), points to a spatially separated chemistry, dominated by oxygen-rich material in the (outer parts of the) circum-binary disk and carbon-rich material in the nebula. There may be a carbon-rich chemistry in the inner disk.

We propose an evolutionary scenario in which the disk is the result of extensive mass transfer and/or mass loss while the central star was still oxygen-rich. The high mass loss exposed carbon-rich interior layers, and the now carbon-rich outflow was forced to a strongly bipolar geometry by the presence of the massive circum-binary disk. Such a model can explain the carbon-rich nature of the extended nebula, the oxygen-rich nature of the (outer parts of the) circum-binary disk and the pronounced X-shape of the nebula. The Red Rectangle may be related to the class of carbon-rich AGB stars with oxygen-rich dust shells²⁴. The oxygen-rich dust near these stars is a remnant of a previous mass-loss episode when the carbon star was still oxygen-rich. The geometry of the oxygen-rich dust shell in these stars is not clear. It may be in a detached, expanding shell²⁴ or in a disk surrounding an unseen companion²⁵ or surrounding the

entire system. The mixed chemistry of the Red Rectangle nebula suggests that a binary model for carbon-stars with oxygen-rich dust shells is plausible. This implies that such binaries may evolve into systems similar to the Red Rectangle.

The strength of the crystalline bands is large compared to that seen in mass-losing AGB stars, but comparable to that seen in some planetary nebulae¹⁷ and in young stars surrounded by proto-planetary systems⁷. The grains in the circum-binary disk are significantly larger than those typical in AGB outflows or the ISM^{22,26}, and the weak CO rotational line emission²⁷ suggests that CO is depleted from the circumstellar environment. The radius of the disk derived from the ISO spectrum corresponds to keplerian velocities of the order of 1 km s^{-1} , which is in good agreement with the width of the CO rotational line emission peak²⁷. This velocity is much smaller than usually found in the outflows of AGB winds ($\sim 10 \text{ km s}^{-1}$), suggesting that the CO gas, and also the dust, are in a stationary circum-binary disk. The grain size distribution, the degree of crystallization, the disk dimensions and total mass, the radial density distribution of the grains as well as the CO kinematics and depletion are reminiscent of the properties of gas and dust in disks surrounding young pre-main-sequence stars^{7,28}. The physical processes that take place in such disks (grain coagulation, crystallization and growth, and—potentially—planet formation) may therefore also occur in the disk of HD44179.

The question arises whether the Red Rectangle may in fact be a young object, as was suggested at its discovery¹. A convincing argument for the evolved nature of the Red Rectangle follows from its membership in a small group of low-gravity objects with similar, extreme depletion of refractory elements in the photosphere², all of which are binaries with orbital periods of about 1 year. The high galactic latitude of the other members of this group strongly argues for an evolved, low-mass nature of the Red Rectangle system. Planet formation so far has not been documented in the post-main-sequence phase of low-mass stars. Nevertheless, because planet formation seems possible in such

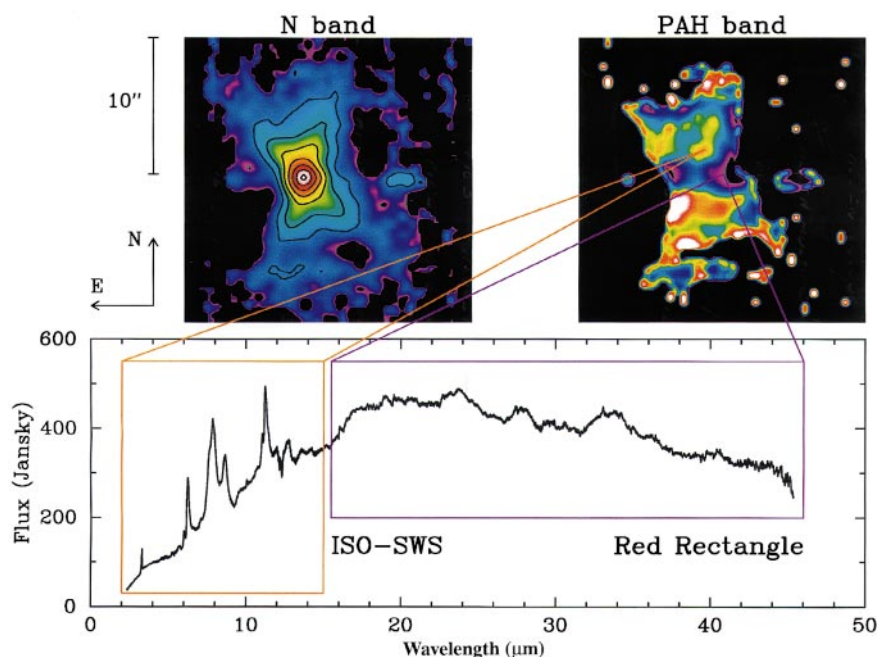


Figure 3 Spatial distribution of the oxygen and carbon rich components of the Red Rectangle. Top, False-colour images of the broad-band 10 μm emission (left) and continuum subtracted narrow-band 11.3 μm emission (right) in the Red Rectangle nebula. The images were taken on 24 February, 1994 with the 10 μm camera TIMMI attached to the 3.6 μm telescope of the European Southern Observatory (ESO), La Silla, Chile. The pixel size is 0.33 arcsec. The broad-band 10 μm image shows that the bulk of the emission at that wavelength originates

from the circum-binary disk, and the brightness distribution of the narrow-band 11.3 μm image shows that the carbon-rich carriers of the UIR bands are located in the extended nebula^{5,6}. Bottom, ISO-SWS spectrum of the Red Rectangle. The boxes relate the carbon-rich components to the extended nebula, and the oxygen-rich component to the circum-binary disk in the centre of the nebula, respectively.

hostile environments as found near millisecond pulsars²⁹, it could well be that planet formation is now occurring in the stationary circumbinary disk of the evolved system HD44179.

There is direct evidence for the presence of magnesium-rich crystalline silicates in the proto-solar nebula, because crystalline enstatite (a magnesium-rich pyroxene) has been found in interplanetary dust particles³⁰. It may very well be that these grains were produced in the outflows of evolved red giants and supergiants. However, so far no strong evidence for the presence of crystalline silicates in the ISM has been found. This implies that crystalline silicates may not be produced sufficiently abundant to be detected in the ISM, where they are mixed with other dust components. Alternatively, the crystalline grains may be destroyed in the ISM. The relatively high abundance of crystalline silicates observed in some proto-planetary disks⁷ suggests that crystallization may also occur during the star formation and planet formation process. □

Received 3 November; accepted 14 December 1997.

1. Cohen, M. *et al.* The peculiar object HD44179 ("The Red Rectangle"). *Astrophys. J.* **196**, 179–189 (1975).
2. van Winckel, H., Waelkens, C. & Waters, L. B. F. M. The extremely iron-deficient "post-AGB" stars are binaries. *Astron. Astrophys.* **293**, L25–L28 (1995).
3. Roddier, F., Roddier, C., Graves, J. E. & Northcott, M. J. Adaptive optics imaging of proto-planetary nebulae: Frosty Leo and the Red Rectangle. *Astrophys. J.* **443**, 249–260 (1995).
4. Osterbart, R., Langer, N. & Weigelt, G. High resolution imaging of the bipolar nebula Red Rectangle. Evidence for unstable mass transfer in a close binary system. *Astron. Astrophys.* **325**, 609–612 (1997).
5. Bregman, J. D., Rank, D., Temi, P., Hudgins, D. & Kay, L. 3.3 and 11.3 micron images of HD44179—evidence for an optically thick polycyclic aromatic hydrocarbon disk. *Astrophys. J.* **411**, 794–796 (1993).
6. Sloan, G. C., Grasdalen, G. L. & Levan, P. D. Spatially resolved spectra of the unidentified infrared features around HD44179 (the Red Rectangle). *Astrophys. J.* **409**, 412–416 (1993).
7. Waelkens, C. *et al.* SWS observations of young main sequence stars with dusty circumstellar disks. *Astron. Astrophys.* **315**, L245–L248 (1996).
8. Russel, R. W., Soifer, B. T. & Willner, S. P. The infrared spectra of CRL618 and HD44179 (CRL915). *Astrophys. J.* **220**, 568–572 (1978).
9. Leger, A. & Puget, J. L. Identification of the 'unidentified' IR emission features of interstellar dust? *Astron. Astrophys.* **137**, L5–L8 (1984).
10. Waelkens, C., van Winckel, H., Waters, L. B. F. M. & Bakker, E. J. Variability and nature of the Red Rectangle nebula. *Astron. Astrophys.* **314**, L17–L20 (1996).
11. Bond, H. E. Chemical compositions of post-AGB stars. in *Evolution of Stars: the Photospheric Abundance Connection* (eds Michaud, G. & Tutukov, A. V.) 341–349 (IAU Symp. 145, Kluwer, Dordrecht, 1991).
12. van Winckel, H., Mathis, J. & Waelkens, C. Evidence from zinc abundances for dust fractionation in chemically peculiar stars. *Nature* **356**, 500–501 (1992).
13. Artymowicz, P. & Lubow, S. H. Mass flow through gaps in circumbinary disks. *Astrophys. J.* **467**, L77–L80 (1996).
14. De Graauw, Th. *et al.* Observing with the ISO Short-Wavelength Spectrometer. *Astron. Astrophys.* **315**, L49–L54 (1996).
15. Kessler, M. F. *et al.* The Infrared Space Observatory (ISO) mission. *Astron. Astrophys.* **315**, L27–L31 (1996).
16. Beintema, D. A. *et al.* The rich spectrum of circumstellar PAHs. *Astron. Astrophys.* **315**, L369–L372 (1996).
17. Waters, L. B. F. M. *et al.* Mineralogy of oxygen-rich dust shells. *Astron. Astrophys.* **315**, L361–L364 (1996).
18. Koike, C., Shibai, H. & Tsuchiyama, A. Extinction of olivine and pyroxene in the mid- and far-infrared. *Mon. Not. R. Astron. Soc.* **264**, 654–658 (1993).
19. Jäger, C., Mutschke, H., Begemann, B., Dorschner, J. & Henning, Th. Steps toward interstellar silicate mineralogy. I: Laboratory results of a silicate glass of mean cosmic composition. *Astron. Astrophys.* **292**, 635–641 (1994).
20. Omont, A. *et al.* Observations of 40–70 micron bands of ice in IRAS09371 + 1212 and other stars. *Astrophys. J.* **355**, L27–L30 (1990).
21. van der Veen, W. E. C. J., Waters, L. B. F. M., Trams, N. R. & Matthews, H. E. A study of dust shells around high latitude supergiants. *Astron. Astrophys.* **285**, 551–564 (1994).
22. Jura, M., Turner, J. & Balm, S. P. Big grains in the Red Rectangle? *Astrophys. J.* **474**, 741–748 (1997).
23. Reese, M. D. & Sitko, M. L. Ultraviolet and visible spectropolarimetry of the Red Rectangle and HD44179. *Astrophys. J.* **467**, L105–L108 (1996).
24. Willems, F. J. & de Jong, T. Carbon stars with oxygen-rich dust shells: Evidence for the onset of the carbon star phase. *Astrophys. J.* **309**, L39–L42 (1986).
25. Morris, M. Mechanisms for mass loss from cool stars. *Publ. Astron. Soc. Pacif.* **99**, 1115–1122 (1987).
26. Lopez, B., Tessier, E., Cruzalebes, P., Lefevre, J. & Le Berre, T. Dust envelope modelling of the Red Rectangle nebula. *Astron. Astrophys.* **322**, 868–877 (1997).
27. Jura, M., Balm, S. P. & Kahane, C. A long-lived disk around the Red Rectangle? *Astrophys. J.* **453**, 721–726 (1995).
28. Mannings, V., Koerner, D. W. & Sargent, A. I. A rotating disk of gas and dust around a young counterpart to β Pictoris. *Nature* **388**, 555–557 (1997).
29. Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature* **355**, 145–147 (1992).
30. Bradley, J. P., Brownlee, D. E. & Veblen, D. R. Pyroxene whiskers and platelets in interplanetary dust: evidence of vapour phase growth. *Nature* **301**, 473–477 (1983).

Acknowledgements. We thank H. Dorschner for providing the laboratory spectrum of olivine. L.B.F.M.W., A.d.K. and J.B. acknowledge financial support from an NWO 'Pionier' grant. F.J.M. acknowledges support from an NWO project grant. C.W. acknowledges financial support from the Belgian Federal Services for Scientific, Technological and Cultural Affairs, and from the Onderzoeksfonds K.U. Leuven. H.V.W. acknowledges financial support from the Fund for Scientific Research of Flanders. We thank A. Boogert for discussions on the CO₂ bands in the Red Rectangle.

Correspondence and requests for materials to L.B.F.M.W. (e-mail: rens@astro.uva.nl).

Dephasing in electron interference by a 'which-path' detector

E. Buks, R. Schuster, M. Heiblum, D. Mahalu & V. Umansky

Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

Wave-particle duality, as manifest in the two-slit experiment, provides perhaps the most vivid illustration of Bohr's complementarity principle: wave-like behaviour (interference) occurs only when the different possible paths a particle can take are indistinguishable, even in principle¹. The introduction of a which-path (*welcher Weg*) detector for determining the actual path taken by the particle inevitably involved coupling the particle to a measuring environment, which in turn results in dephasing (suppression of interference). In other words, simultaneous observations of wave and particle behaviour is prohibited. Such a manifestation of the complementarity principle was demonstrated recently using a pair of correlated photons, with measurement of one photon being used to determine the path taken by the other and so prevent single-photon interference². Here we report the dephasing effects of a which-path detector on electrons traversing a double-path interferometer. We find that by varying the sensitivity of the detector we can affect the visibility of the oscillatory interference signal, thereby verifying the complementarity principle for fermions.

Mesoscopic systems³ can be used to study the interplay between interference and dephasing of electrons⁴. Nano-scale fabrication and low-temperature measuring techniques, which minimize unintentional dephasing, enable the observation of variety of coherent effects of electrons such as an induced Aharonov–Bohm phase, weak localization, resonant tunnelling and conductance quantization⁵. A controllable dephasing via a which-path detector, on the other hand, has not been previously demonstrated in such systems.

In our experiment we used a double-path electronic interferometer⁶, fabricated within the plane of a high-mobility two-dimensional electron gas. The two paths are defined by two slits electrons can pass through, with one slit in the form of a coherent quantum dot (QD)^{5,6}. The QD is a trap that captures electrons for a relatively long time, like a resonant delay line, thus allowing the electrons to be detected more easily. Near the QD, but electrically separated from it, a quantum point contact (QPC) is fabricated, serving as a which-path detector. The QPC is a short conducting segment with width comparable to the electron wavelength, allowing only a small number of modes to pass. It is expected that an electron passing the QD-slit will interact with the nearby QPC-detector (both systems are thus 'entangled'⁷) and modify the conductance of the QPC⁸. This detection process leads to dephasing, that is, to a suppression of the double-path Aharonov–Bohm interference.

The two-path interferometer (Fig. 1) consists of emitter E and collector C constrictions; each constriction is made by a single-mode QPC, and there is a base region B between them. The grounded base contacts (with voltage $V_B = 0$) serve as draining reservoirs for back-scattered electrons, ensuring that only two forward-propagating paths reach the collector. The emitter is separated from the collector by a barrier with two slits; one of those slits is a QD. As the collector signal is small, we incorporated additional reflecting barriers (the white gates in Fig. 1a) to direct the emitted electrons into the two slits and subsequently to reflect them towards the collector. All measurements were done at an electrons

temperature $\Theta \approx 80$ mK and an a.c. emitter excitation voltage $V_E = 10 \mu\text{V}$. Under these conditions both coherence length (due to unintentional dephasing) and mean free path for elastic scattering of the electrons exceed the entire size of the interferometer.

The collector current is related to the transmission probability from emitter to collector, T_{EC} , via the multiprobe conductance formula⁹, $I_C = (2e^2/h)T_{EC}V_E$. As stressed above, the dominant contribution to T_{EC} comes from the two direct paths, those going from E to C through the two slits (depicted by the two dotted lines in Fig. 1a; longer paths, resulting from multiple reflections from walls, are much less probable. A phase difference between the two direct paths, $\Delta\alpha = 2\pi\Phi/\Phi_0$, is induced via the Aharonov–Bohm effect (for a review see ref. 10). Here Φ is the magnetic flux threaded through the area, A , enclosed by these two paths and $\Phi_0 = h/e$ is the flux quantum. Consequently, the collector current oscillates as a function of magnetic field B with a period $\Delta B = \Phi_0/A = 2.6$ mT, corresponding to a phase difference between the two paths equal to 2π , as seen in Fig. 2a.

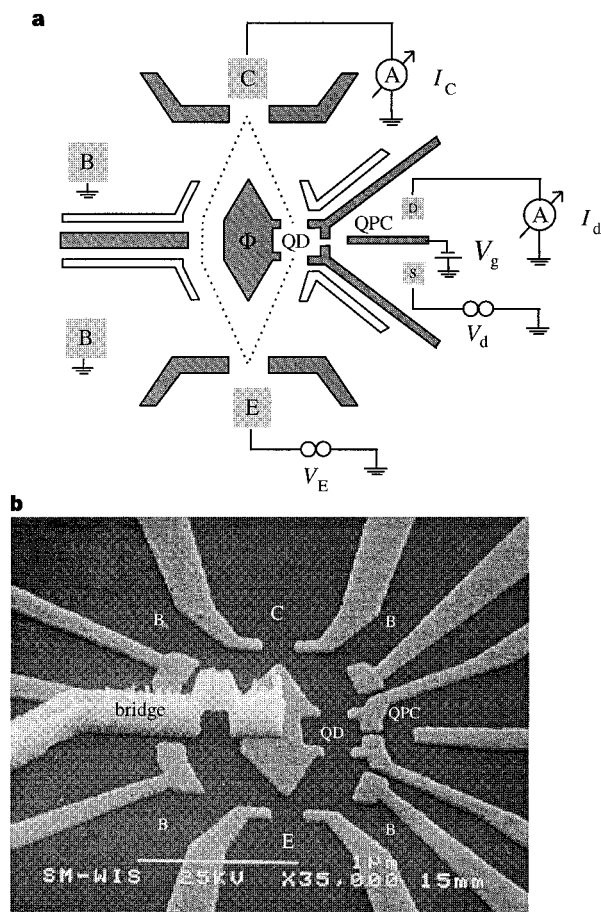


Figure 1 The which-path device. **a**, Schematic description of the top electrodes and contacts of the interferometer and the detector. The interferometer is composed of three different regions, emitter E, collector C, and base B. The right slit is in a form of a quantum dot, QD (with area $0.4 \times 0.4 \mu\text{m}^2$) with a quantum point contact, QPC, on its right side serving as a which-path detector. The excitation voltage V_E is applied between emitter and base. **b**, Top-view scanning electron micrograph of the device. Scale bar, $1 \mu\text{m}$. The GaAs–AlGaAs heterostructure supports a high-mobility two-dimensional electron gas (2DEG) (with density $n_s = 3.0 \times 10^{11} \text{ cm}^{-2}$ and low-temperature mobility $\mu = 2.8 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) formed 60 nm beneath the surface. Potential barriers and narrow openings in the plane of the 2DEG are induced by negatively biased miniature metal gates deposited on the surface of the heterostructure, thus depleting the electrons underneath the gates. A special lithographic technique, involving a metallic air bridge, is used to contact the central gate that depletes the area between the two slits (this serves also as plunger gate of the QD).

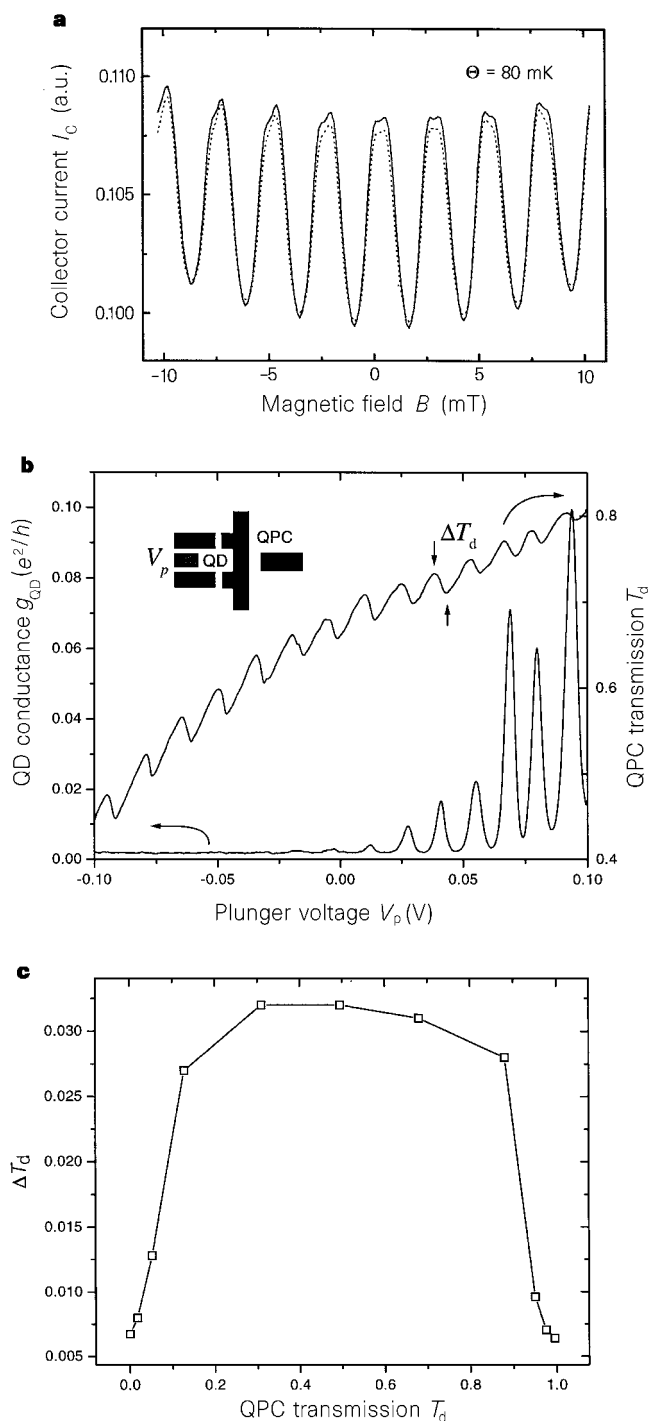


Figure 2 Conduction characteristics of the which-path device. **a**, Aharonov–Bohm oscillations of the collector current I_C with a period $\Delta B = \Phi_0/A = 2.6$ mT. The solid line is measured with QPC drain source voltage $V_d = 0$ (which-path detector turned off), while the dotted line, with a reduced visibility, is measured with $V_d = 100 \mu\text{eV}$. (a.u., arbitrary units.) **b**, The conductance of the QD, g_{QD} , and the transmission of the QPC nearby, T_d , as a function of the plunger gate voltage, V_p . The inset shows schematically the coupled structures. **c**, The induced average change in the transmission probability of the QPC detector, ΔT_d , due to adding an electron to the QD as a function of T_d . ΔT_d is calculated by averaging over several coulomb-blockage peaks. The reduced value of ΔT_d near $T_d = 0$ and $T_d = 1$ is a consequence of approaching the conductance plateaus.

The suppression of Aharonov–Bohm oscillations due to the which-path detector depends on the effect an electron dwelling in the QD has on the detector. We study the QD–QPC interaction by means of a calibration device (seen in Fig. 2b inset), fabricated on the same wafer, containing a QD and a QPC similar to these in the which-path interferometer. The QD is tuned to the coulomb blockade regime, having therefore well separated energy levels, by adjusting the resistance of each of its two QPCs to be greater than $\hbar/2e^2$. Each peak in the conductance of the QD, scanned by the plunger gate voltage V_p (see Fig. 2b), is associated with adding a single electron in the QD. Owing to the proximity between the QD and QPC, the transmission probability of the QPC-detector T_d , and thus its conductance $g_d = (2e^2/h)T_d$, are affected by the potential of the QD⁸. As the plunger gate voltage is being scanned between two adjacent conductance peaks, with fixed charge in the QD, the potential of the QD changes smoothly. However, when a conductance peak is being scanned, with an electron being added to the QD, a faster and opposite change in the potential of the QD takes place, on the scale of the peak width. Consequently, the potential of the QPC, and therefore T_d , are expected to show a saw-tooth-like oscillations (with amplitude ΔT_d), as indeed observed in the experimental results in Fig. 2b (see also ref. 8).

We now consider the expected quantitative dephasing induced by the WP detector. This has been treated recently by Aleiner *et al.*¹¹, Levinson¹², Gurvitz¹³, and Imry¹⁴. Following ref. 4, we write the entangled wavefunction of the whole system (interferometer + detector) as:

$$|\psi\rangle = |\phi_l\rangle_e \otimes |\chi_l\rangle_d + e^{i\Delta\alpha} |\phi_r\rangle_e \otimes |\chi_r\rangle_d \quad (1)$$

where $|\phi_l\rangle_e$ ($|\phi_r\rangle_e$) is the electronic partial wave in the left (right) path, and $|\chi_l\rangle_d$ ($|\chi_r\rangle_d$) represents the state of the detector coupled to the left (right) partial electronic wave. The probability to find the electron at the collector is found by summing over all possible states of the detector (as the collector is sensitive only to the electron position, regardless the state of the detector), $T_{EC} = \sum_i |\langle\psi|\mathbf{r}_C\rangle_e \otimes |\chi_i\rangle_d|^2$ where $|\mathbf{r}_C\rangle_e$ represents the state of an electron at the collector. Assuming $|\chi_l\rangle_d$ and $|\chi_r\rangle_d$ are normalized one finds:

$$T_{EC} = |\langle\phi_l|\mathbf{r}_C\rangle_e|^2 + |\langle\phi_r|\mathbf{r}_C\rangle_e|^2 + 2\text{Re}[e^{i\Delta\alpha} \langle\phi_l|\mathbf{r}_C\rangle_e \cdot \langle\mathbf{r}_C|\phi_r\rangle_e \cdot \langle\chi_r|\chi_l\rangle_d] \quad (2)$$

(where Re denotes real part) with the visibility of the interference pattern, defined as the peak-to-valley value normalized by the average value, being proportional to:

$$v_d = |\langle\chi_r|\chi_l\rangle_d| \quad (3)$$

Here v_d represents which-path information that can be obtained from the detector¹⁵, namely, the smaller v_d is, the stronger is the dephasing and the weaker is the visibility. In the extreme case, when the detector determines unambiguously where the electron is, both detector states $|\chi_l\rangle_d$ and $|\chi_r\rangle_d$ are orthogonal and the Aharonov–Bohm interference vanishes.

How certain is our which-path detection? An electron entering the QD-slit changes the transmission probability of the QPC-detector by ΔT_d . The rate at which particles probe the detector at zero temperature is $2eV_d/h$, where V_d is the voltage across the detector. Thus, the number of particles probing the detector during the dwell time of the electron in the QD, $\tau_d = \hbar/2\pi\Gamma$ (Γ is the elastic width of the resonant state), is $N = \tau_d 2eV_d/h = (1/\pi)eV_d/\Gamma$. Now, we let N_t be the number of transmitted particles out of the total N particles probing the detector. This is a binomial random variable (leading to ‘shot noise’ in the current), having an expectation value $\langle N_t \rangle = NT_d$ and a standard deviation $\sigma(N_t) = \sqrt{NT_d(1-T_d)}$ (refs 16, 17). To determine the certitude of the detection and thus the extent of dephasing, ΔT_d has to be compared with the resultant uncertainty in T_d , given by $\sigma(T_d) = \sqrt{T_d(1-T_d)/N}$. For a noisy detector, $\sigma(T_d) \gg \Delta T_d$: the detector does not provide which-path information and one expects distinct interference, namely $v_d \approx 1$. Whereas for a quiet detector, $\sigma(T_d) \ll \Delta T_d$: one can determine, even if ‘in principle’, the path the electron takes, and consequently, the interference pattern is expected to diminish.

A direct evaluation of the overlap v_d in equation (3), for a symmetric barrier in the QPC detector¹¹, leads to $v_d = 1 - (1/8)[\Delta T_d/\sigma(T_d)]^2$ at zero temperature, namely:

$$v_d = 1 - \frac{1}{\pi} \frac{eV_d}{\Gamma} \frac{(\Delta T_d)^2}{8T_d(1-T_d)} \quad (4)$$

We measured the visibility of the Aharonov–Bohm oscillations when the QD-slit is tuned to a conduction peak (using the central

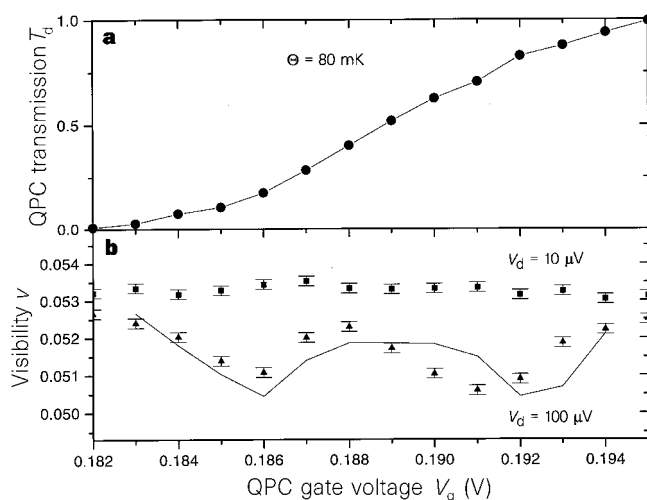
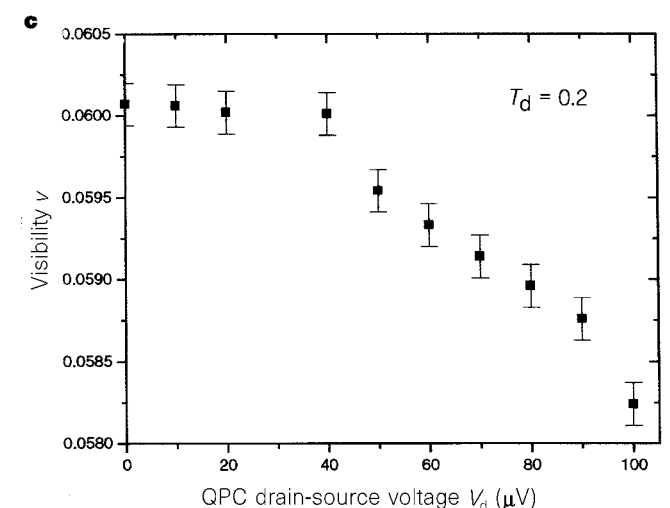


Figure 3 Measurements of visibility. **a**, The transmission probability of the QPC detector, T_d , as a function of the voltage applied to the right gate of the QPC-detector, V_g . **b**, The visibility (v) of the Aharonov–Bohm (AB) oscillations as a function of V_g for two values of drain source voltage, V_d . The peak-to-valley value of the AB oscillations is obtained using the following procedure. First, the non-oscillatory component of the trace I_C versus B is subtracted using a least-mean-squared fit to a polynomial of degree 2. Second, numerical integration of the



remaining oscillatory component squared leads to the peak-to-valley value. Last, the visibility is found by dividing by the average value of I_C . Error bars indicate the fluctuations in visibility due to fluctuations of device’s properties (instrumental noise is negligibly small). **c**, The visibility of the AB conductance oscillations as a function of V_d for a fixed $T_d = 0.2$. The behaviour is linear for $eV_d > k_B\Theta$ with saturation for low V_d .

metal island as a plunger gate), and the QPC-detector conducting current with transmission $0 \leq T_d \leq 1$. Figure 3a shows the transmission probability of the QPC detector, T_d , and Fig. 3b shows the visibility for two values of V_d , both as a function of V_g (see Fig. 1a). For $V_d = 100 \mu\text{V}$ the observed visibility peaks when ΔT_d of the QPC-detector is small near conductance plateaus, and also near $T_d = 0.5$, with the quantum shot noise in the QPC-detector at maximum. In these regimes the which-path detector provides least-certain information about the location of the electron and therefore dephasing is small. These features disappear when V_d is reduced to $10 \mu\text{V}$. This observation confirms that the behaviour of the visibility found for $V_d = 100 \mu\text{V}$ is indeed due to dephasing, and is not related to undesirable electrostatic effect of V_g on the QD.

To compare our experimental results for the case $V_d = 100 \mu\text{V}$ with theory, we use the zero-temperature value of v_d in equation (4) (valid because $eV_d \gg k_B \Theta \approx 7 \mu\text{eV}$ (E.B., unpublished results)). We use the measured ΔT_d in the calibration device (see Fig. 2c) and a value $\Gamma = 0.5 \mu\text{eV}$ (corresponding to $\tau_d \approx 1.3 \text{ ns}$) as a fitting parameter. The calculated visibility, drawn as a solid line in Fig. 3b, shows reasonable agreement with experiment, both qualitatively and quantitatively. The dependence of the visibility on drain source voltage V_d (measured in a different working regime from that shown in Fig. 3b) is given in Fig. 3c. For $eV_d \gg k_B \Theta$, the visibility drops linearly as the probing rate of the QPC-detector by the electrons increases, as expected from equation (4). The deviation from the linear dependence near $V_d = 0$ can be accounted for by the fact that $eV_d \approx k_B \Theta$ (E.B., unpublished results).

Apart from demonstrating the principle of complementarity, we believe that similar experimental set-ups, but with higher detector sensitivity, may be used to study other fundamental problems in quantum mechanics. For example, looking at the current of the which-path detector in the time domain may shed light on the old and controversial issue of wavefunction collapse. Moreover, increasing the mutual coupling between detector and interferometer might open a way to fabrication of an electronic quantum bit, with possible applications in quantum computing. □

Received 17 September; accepted 16 December 1997.

- Bohr, N. in *Albert Einstein: Philosopher—Scientist* (ed. Schilpp, P. A.) 200–241 (Library of Living Philosophers, Evanston, 1949).
- Zou, X. Y., Wang, L. J. & Mandel, L. Induced coherence and indistinguishability in optical interference. *Phys. Rev. Lett.* **67**, 318–321 (1991).
- Imry, Y. *Introduction to Mesoscopic Physics* (Oxford Univ. Press, 1997).
- Stern, A., Aharonov, Y. & Imry, Y. Phase uncertainty and loss of interference: a general picture. *Phys. Rev. A* **41**, 3436–3448 (1990).
- Schuster, R. *et al.* Phase measurement in a quantum dot via a double-slit interference experiment. *Nature* **385**, 417–420 (1997).
- Van Houten, H., Beenakker, C. W. J. & Staring, A. A. W. in *Single Charge Tunneling—Coulomb Blockade Phenomena in Nanostructures* (eds Grabert, H. & Devoret, M. H.) Ch. 5 (Plenum, New York, 1992).
- Gurvitz, S. A. Interaction free measurement in mesoscopic systems and the reduction postulate. Preprint available at <http://xxx.lanl.gov/abs/quant-ph/9607029>.
- Field, M. *et al.* Measurements of Coulomb blockade with a noninvasive voltage probe. *Phys. Rev. Lett.* **70**, 1311–1314 (1993).
- Buttiker, M. Four terminal phase coherent conductance. *Phys. Rev. Lett.* **57**, 1761–1764 (1986).
- Aronov, A. G. & Sharvin, Yu. V. Magnetic flux effects in disordered conductors. *Rev. Mod. Phys.* **59**, 755–779 (1987).
- Aleiner, I. L., Wingreen, N. S. & Meir, Y. Dephasing and the orthogonality catastrophe in tunneling through a quantum dot: the “which path” interferometer. *Phys. Rev. Lett.* **79**, 3740–3743 (1997).
- Levinson, Y. Dephasing in a quantum dot due to coupling with a quantum point contact. *Europhys. Lett.* **39**, 299–304 (1997).
- Gurvitz, S. A. Measurements with a noninvasive detector and dephasing mechanism. *Phys. Rev. B* **56**, 15215–15223 (1997).
- Imry, Y. in *Proc. 1997 Nobel Symp. on Modern Studies of Basic Quantum Concepts and Phenomena* (in the press).
- Egler, B. G. Fringe visibility and which-way information: an inequality. *Phys. Rev. Lett.* **77**, 2154–2157 (1996).
- Khlos, V. A. Current and voltage fluctuations in microjunctions between normal metals and superconductors. *JETP* **66**, 1243–1249 (1987).
- Lesovik, G. B. Excess quantum noise in 2D ballistic point contacts. *JETP Lett.* **49**, 592–594 (1989).

Acknowledgements. We thank S. Gurvitz for presenting to us ref. 7, which initiated the present work. We also thank I. Imry, Y. Levinson, Y. Meir, A. Stern and N. Wingreen for discussions. This work was supported in part by a MINERVA grant and a MINERVA fellowship for one of us (R.S.).

Correspondence and requests for materials should be addressed to M.H. (e-mail address: heiblum@wis.weizmann.ac.il).

Two-dimensional ferroelectric films

A. V. Bune[†], V. M. Fridkin[†], Stephen Ducharme^{*}, L. M. Blinov[†], S. P. Palto[†], A. V. Sorokin[†], S. G. Yudin[†] & A. Zlatkin[†]

^{*} Department of Physics and Astronomy and the Center for Materials Research and Analysis, University of Nebraska, Lincoln, Nebraska 68588-0111, USA

[†] Institute of Crystallography of the Russian Academy of Sciences, 117 333 Moscow, Russia

Ultrathin crystalline films offer the possibility of exploring phase transitions in the crossover region between two and three dimensions. Second-order ferromagnetic phase transitions have been observed in monolayer magnetic films^{1,2}, where surface anisotropy energy stabilizes the two-dimensional ferromagnetic state at finite temperature³. Similarly, a number of magnetic materials have magnetic surface layers that show a second-order ferromagnetic–paramagnetic phase transition with an increased Curie temperature⁴. Ferroelectricity is in many ways analogous to ferromagnetism, and bulk-like ferroelectricity and finite-size modifications of it have been seen in nanocrystals as small as 250 Å in diameter⁵, in perovskite films 100 Å thick⁶ and in crystalline ferroelectric polymers as thin as 25 Å (refs 7–10). But these results can be interpreted as bulk ferroelectricity suppressed by surface depolarization energies, and imply that the bulk transition has a minimum critical size^{11–13}. Here we report measurements of the ferroelectric transition in crystalline films of a random copolymer of vinylidene fluoride and trifluoroethylene just 10 Å (two monolayers) thick. We see a first-order ferroelectric phase transition with a transition temperature nearly equal to the bulk value, even in these almost two-dimensional films. In addition, we see a second first-order transition at a lower temperature, which seems to be associated with the surface layers only. The near-absence of finite-size effects on the bulk transition implies that these films must be considered as two-dimensional ferroelectrics.

Ferroelectric polymers have been studied in great detail for nearly 30 years and are in wide use in a variety of piezoelectric devices^{14,15}. One of the most interesting systems is the random copolymer of vinylidene fluoride with trifluoroethylene, P(VDF-TrFE) (ref. 16), consisting of $-(\text{CF}_2\text{—CH}_2)_x\text{—}(\text{CF}_2\text{—CHF—})_{1-x}\text{—}$ chains with a regular intra-chain period of 2.6 Å, controlled by the C—C bonds between fluorine pairs, as shown in Fig. 1a. In the ferroelectric phase, the all-*trans* chains are arranged in parallel rows in a quasi-hexagonal close packing with orthorhombic *mm2* structure¹⁷. P(VDF-TrFE 70:30) has a first-order ferroelectric–paraelectric phase transition (connected with the conversion of all-*trans* chains to mixtures of *trans* and *gauche* bonds with little or no net dipole moment¹⁶) with Curie temperature $T_c \approx 100^\circ\text{C}$ (the Curie temperature increases with the proportion of the VDF component), thermal hysteresis of $\Delta T_c \approx 20^\circ\text{C}$, and a spontaneous polarization $P_s \approx 0.1 \text{ C m}^{-2}$ at room temperature¹⁶. Thin films have usually been formed by solvent spinning and are polymorphous, containing amorphous material and incompletely orientated crystallites. Finite-size effects were evident from an increase in the coercive field in films as thin as 600 Å (ref. 18).

The first crystalline Langmuir–Blodgett-deposited polymer films, of P(VDF-TrFE 70:30), achieved a great improvement in quality^{7,8,10}. They have excellent crystalline order and show the first-order ferroelectric–paraelectric transition at $T_c \approx 80^\circ\text{C}$ on cooling. These ferroelectric polymer films, 30 monolayers (ML) thick ($\sim 150 \text{ Å}$), also showed double hysteresis and the critical point¹⁹, and a new conductance switching controlled by the polarization state⁸. Finite-size effects have been demonstrated in these films; the

coercive field scaled as the -0.7 power of the thickness in films with 35–135 ML (ref. 9).

For the present work we first recorded an atomic-resolution image of a single-monolayer P(VDF-TrFE 70:30) film on graphite by scanning tunnelling microscopy (STM, Fig. 1b) showing the excellent short-range order and uniform orientation of all the all-*trans* polymer chains. Figure 2a–c shows that multilayer ferroelectric Langmuir–Blodgett-deposited polymer (FLP) films with aluminium electrodes have the dielectric anomalies and thermal hysteresis that mark the usual bulk first-order ferroelectric phase transition found at 77°C on cooling (108°C on heating) in the 30-ML film and decreasing to 68°C (98°C) in the 2-ML film. The pyroelectric response shown in Fig. 3 also records the bulk transition near 73°C , though the precise transition temperature is not as clear as in the dielectric measurements. Therefore, the ‘bulk’ first-order phase transition at $\sim 80^\circ\text{C}$ remains even in the 2-ML films (Fig. 2c)—a result that challenges theoretical predictions that the transition temperature should decrease, and ferroelectricity vanish, in films below a minimum critical thickness^{11–13}. FLP films of 30 ML or less all have approximately the same room-temperature relative dielectric constant $\epsilon = 8 \pm 2$, as shown in Fig. 2d, approximately one-half the value found in thick spun films.

The hysteresis loops shown in Fig. 4 show good saturation for the 5-ML and 30-ML films and incomplete saturation for the 2-ML film. The 5-ML and 2-ML films show considerable bias, probably due to interactions with the substrate or the top electrode, consistent with our earlier observations on the dynamics of switching FLP films^{8,9}. The coercive fields of $0.9 \pm 0.2 \text{ GV m}^{-1}$ (30 ML), $0.6 \pm 0.1 \text{ GV m}^{-1}$ (5 ML) and $>0.5 \text{ GV m}^{-1}$ (2 ML) are far larger than in other ferroelectric materials and are closely approaching the intrinsic coercive field $P_s/\epsilon\epsilon_0 \approx 1 \text{ GV m}^{-1}$ expected in the absence of nucleation, consistent with the very high crystallinity (here ϵ_0 is the permittivity of free space). The films achieved at least 50% of the

maximum polarization expected if all dipoles are aligned perpendicular to the film, but our data could not rule out a tendency towards in-plane polarization as is observed in ultrathin magnetic films²⁰.

Figures 2 and 3 also show a new and unexpected phase transition at a lower temperature of $\sim 20^\circ\text{C}$ in films of 30 ML or less. The dielectric constant (Fig. 2) shows peaks at $\sim 20^\circ\text{C}$ on cooling (28°C on heating) in the pyroelectric response (Fig. 3) in the 5-ML and 2-ML films. The polarization hysteresis loops from films of 30, 5 and 2 ML (Fig. 4) show that ferroelectric switching is obtained at 25°C in all the FLP films, even those as thin as 2 ML. We attribute the behaviour at $\sim 20^\circ\text{C}$ to a new first-order ferroelectric transition in the surface layers on the basis of several observations: the new transition is strongest in the thinnest films; there is clear thermal hysteresis in the dielectric constant; and there are peaks in both the dielectric constant and the pyroelectric response. There is other evidence that the surface layers behave differently to the interior layers, because most films of 30 ML or less are found to be partially or fully polarized immediately after fabrication, even before the application of an external electric field. Also, the polarization of the 2–7 ML films reverts in zero field to an orientation determined by the substrate, independent of the history of electric-field

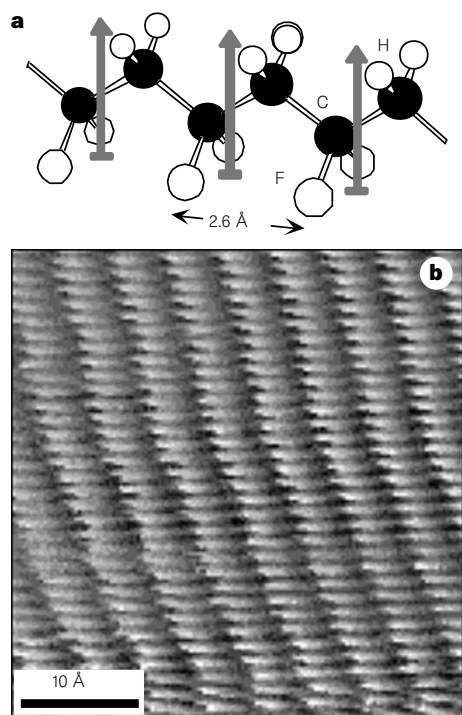


Figure 1 Structure of the polymer chains and films. **a**, Fragment of a P(VDF-TrFE 70:30) copolymer chain. 30% of the VDF ($-\text{CH}_2-\text{CF}_2-$) units have been replaced at random by TrFE ($-\text{CHF}-\text{CF}_2-$) units. The arrows show the direction of the net dipole moments pointing from the fluorine side to the hydrogen side of the chain. **b**, Atomic-resolution STM image of a single-monolayer film of P(VDF-TrFE 70:30) deposited on graphite.

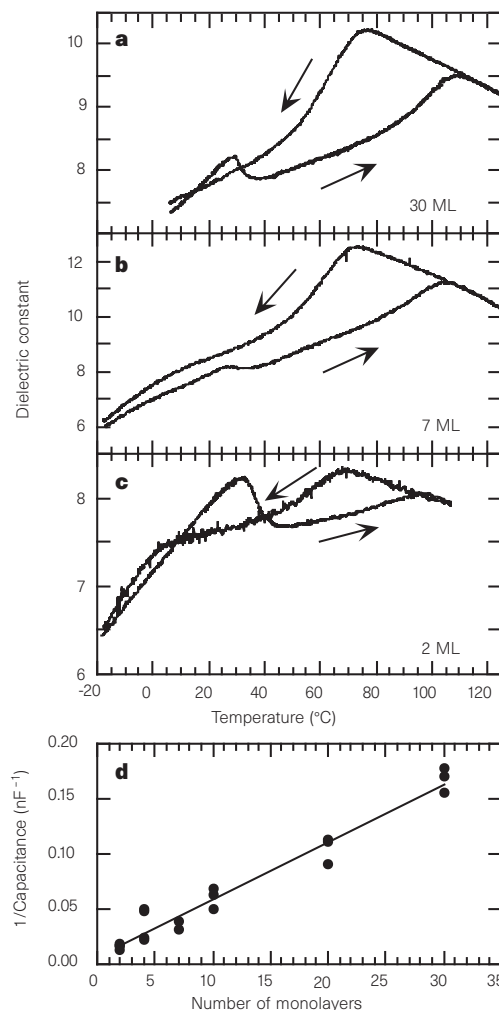


Figure 2 Dielectric properties of the FLP films. **a–c**, Dielectric constant ϵ of the P(VDF-TrFE 70:30) films with different thicknesses, showing peaks at the phase transitions on heating and, due to thermal hysteresis, at different temperatures on cooling: **a**, 30 ML; **b**, 7 ML; **c**, 2 ML. The arrows show the direction of temperature change. **d**, The inverse capacitance at 25°C for FLP films with different thicknesses.

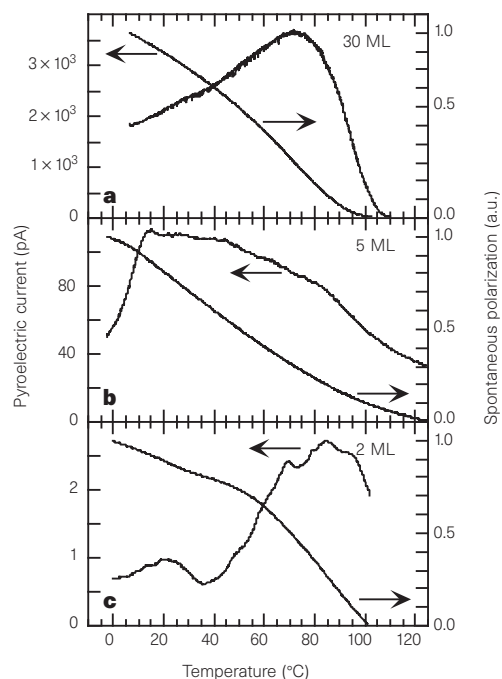


Figure 3 Pyroelectric response and spontaneous polarization P_s , obtained by integration over temperature, of P(VDF-TrFE 70:30) films. **a**, 30 ML; **b**, 5 ML; **c**, 2 ML.

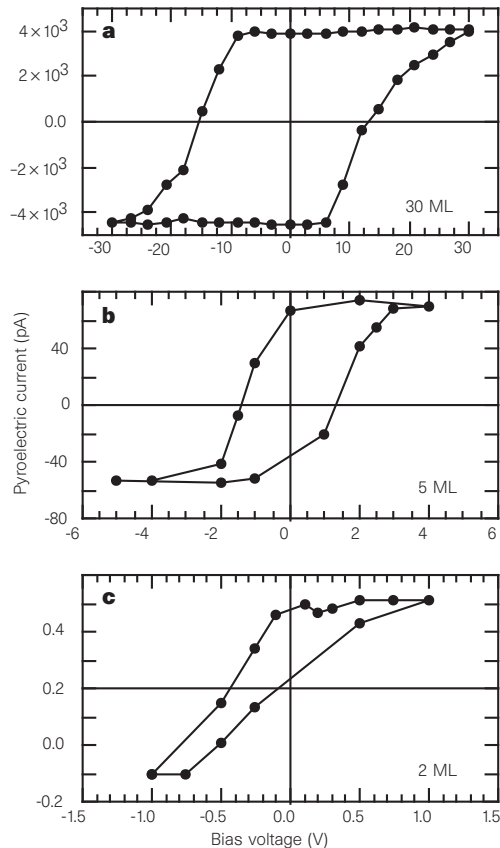


Figure 4 Polarization hysteresis loops at 25°C, measured by the pyroelectric technique, of the P(VDF-TrFE 70:30) films with different thicknesses. **a**, 30 ML; **b**, 5 ML; **c**, 2 ML.

application^{8,9}, consistent with the failure of the pyroelectric response to completely vanish above the Curie temperature in the thinnest films. These observations indicate that the surface layers have a preferred direction controlled by the interaction with the substrate or top electrode.

The present results lead to the following conclusions. First, there is no apparent critical (minimum) thickness in FLP films as thin as 10 Å, much thinner than any previous ferroelectric films, with only a small suppression of the bulk Curie temperature. Second, there is a surface-layer ferroelectric phase transition at ~20°C observable only in the thinnest FLP films. Therefore, the FLP films must be considered essentially two-dimensional ferroelectrics because of the absence of strong finite-size effects; neither the coercive field nor the bulk phase transition temperature change significantly as the film thickness decreases from 150 to 10 Å, implying that the interlayer coupling is weak. Finite-size calculations based on a three-dimensional mean-field theory are not appropriate for modelling the thinnest FLP films with thickness comparable to size of the individual dipoles. The mean-field theory also predicts the existence of a critical thickness equal to the ferroelectric correlation length (a phenomenological parameter that is unknown in ferroelectric polymers²¹); the surface energy, connected with the truncation of the crystal and the dielectric properties of the external material, introduces another length scale. Similar considerations have been used to calculate²² a critical size ~25 Å for small spherical particles of lead zirconate-titanate (PZT).

The two-dimensional nature of the FLP films means that the ferroelectric state may be generated by coupling only within the plane of the film. Any coupling between planes is weak and possibly responsible for the lower surface-layer Curie temperature. The Ising model for ultrathin films^{23,26} is a more appealing approach to modelling ferroelectricity in two-dimensional polymer films because the dipole moments have restricted freedom—they can rotate only about the chain axis and are further inhibited from rotations about the axis by both interchain steric interactions and intrachain dihedral stiffness. We expect that an appropriate Ising model could be constructed in either of two ways. First, use of anisotropic coupling constants (like the exchange integrals in ferromagnetism) with strong ferroelectric coupling in the plane and a weak coupling perpendicular to the plane. Second, use of a purely two-dimensional model and making a weak inter-plane coupling through a mean-field shared by all layers. Both approaches can achieve ferroelectricity at finite temperature in a single layer and either enhancement or suppression by neighbouring layers or electrodes, depending on the sign of the interlayer coupling. Both approaches also suggest that the surface layers, boundaries between the ferroelectric film and the electrodes or other outside material, have a Curie point different from the interior 'bulk' layers because they couple with only one other ferroelectric layer. Fluctuations in two dimensions are not expected to destroy ordering as in the case of the isotropic Heisenberg ferromagnet³ because the coupling is intrinsically anisotropic, and the polymer dipoles have close to one rotational degree of freedom, compared to the two rotational degrees of freedom of the magnetic dipoles. □

Methods

The crystalline FLP films were formed by horizontal Langmuir–Blodgett deposition from a water subphase, as reported in greater detail previously^{7,24}. The STM image (Fig. 1b) was obtained using an MTD (Moscow, Russia) instrument with a 1-ML FLP film deposited on a graphite substrate²⁵. The films for electric measurements were deposited on aluminium-coated glass substrates and overcoated with aluminium evaporated in vacuum. The dielectric response (Fig. 3) was measured with a commercial RLC meter (Hewlett-Packard 4192A) at 1 kHz frequency and 0.1 V amplitude. The pyroelectric response at zero bias (Fig. 2) and the hysteresis loops were measured by periodically heating the sample, with an 80-mW laser beam operating at 514.5-nm wavelength and modulated at 2 kHz, and measuring the short-circuit current

with a lock-in amplifier. The relative spontaneous polarization was obtained by integrating the pyroelectric current over temperature.

Received 5 June; accepted 28 October 1997.

1. Dürr, W. *et al.* Magnetic phase transition in two-dimensional ultrathin Fe films on Au(100). *Phys. Rev. Lett.* **62**, 206–209 (1989).
2. Farle, M. & Baberschke, K. Ferromagnetic order and the critical exponent γ for a Gd monolayer: an electron-spin-resonance study. *Phys. Rev. Lett.* **58**, 511–514 (1987).
3. Mermin, N. D. & Wagner, H. Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.* **17**, 1133–1136 (1966).
4. Dowben, P. A., McIlroy, D. N. & Li, D. In *Handbook on the Physics and Chemistry of Rare Earths* (eds Gschneidner, K. A. Jr & Eyring, L.) Ch. 159 (Elsevier, Amsterdam, 1997).
5. Ishikawa, K., Yoshikawa, K. & Okada, N. Size effect on the ferroelectric phase transition in PbTiO₃ ultrafine particles. *Phys. Rev. B* **37**, 5852–5855 (1988).
6. Karasawa, J., Sugiura, M., Wada, M., Hafid, M. & Fukami, T. Ultra-thin lead titanate films prepared by tripole magnetron sputtering. *Ferroelectr. Lett.* **12**, 105–114 (1996).
7. Palto, S. *et al.* Ferroelectric Langmuir-Blodgett films. *Ferroelectr. Lett.* **19**, 65–68 (1995).
8. Bune, A. *et al.* Novel switching phenomena in ferroelectric Langmuir-Blodgett films. *Appl. Phys. Lett.* **67**, 3975–3977 (1995).
9. Blinov, L. M., Fridkin, V. M., Palto, S. P., Sorokin, A. V. & Yudin, S. G. Thickness dependence of switching for ferroelectric Langmuir films. *Thin Solid Films* **284–285**, 474–476 (1996).
10. Sorokin, A., Palto, S., Blinov, L., Fridkin, V. & Yudin, S. Ultrathin ferroelectric Langmuir-Blodgett films. *Mol. Mater.* **6**, 61–67 (1996).
11. Scott, J. F. Phase transitions in ferroelectric thin films. *Phase Trans.* **30**, 107–110 (1991).
12. Tilley, D. R. in *Ferroelectric Thin Films: Synthesis and Basic Properties* (eds Paz de Araujo, C., Scott, J. F. & Taylor, G. F.) 11–45 (Gordon & Breach, New York, 1996).
13. Scott, J. F. Properties of ceramic KNO₃ thin film memories. *Physica B* **150**, 160–167 (1988).
14. Lovinger, A. J. Ferroelectric polymers. *Science* **220**, 1115–1121 (1983).
15. Wang, T. T., Herbert, J. M. & Glass, A. M. (eds) *The Applications of Ferroelectric Polymers* (Chapman & Hall, New York, 1988).
16. Furukawa, T. Ferroelectric properties of vinylidene fluoride copolymers. *Phase Transit.* **18**, 143–211 (1989).
17. Legrand, J. F. Structure and ferroelectric properties of P(VDF-TrFE) copolymers. *Ferroelectrics* **91**, 303–317 (1989).
18. Kimura, K. & Ohigashi, H. Polarization behavior in vinylidene fluoride-trifluoroethylene copolymer thin films. *Jpn J. Appl. Phys.* **25**, 383–387 (1986).
19. Ducharme, S. *et al.* Critical point in ferroelectric Langmuir-Blodgett polymer films. *Phys. Rev. B* **57**, 25–28 (1998).
20. Allenspach, R. & Bishof, A. Magnetization direction switching in Fe/Cu(110) epitaxial films: temperature and thickness dependence. *Phys. Rev. Lett.* **69**, 3385–3388 (1992).
21. Qu, B. D., Zhang, P. L., Wang, Y. G., Wang, C. L. & Zhong, W. L. Dielectric susceptibility of ferroelectric thin films. *Ferroelectrics* **152**, 219–224 (1994).
22. Yamamoto, T. Calculated size dependence of ferroelectric properties in PbZrO₃-PbTiO₃ system. *Integr. Ferroelectr.* **12**, 161–166 (1996).
23. Wang, C. L., Zhong, W. L. & Zhang, P. L. The Curie temperature of ultra-thin ferroelectric films. *J. Phys.: Cond. Matter* **3**, 4743–4749 (1992).
24. Blinov, L. M. Langmuir films. *Sov. Phys. Usp.* **31**, 623–644 (1988).
25. Palto, S. *et al.* Ferroelectric Langmuir-Blodgett films showing bistable switching. *Europhys. Lett.* **34**, 465–469 (1996).
26. Zhang, R. & Taylor, P. L. Theory of ferroelectric-paraelectric transitions in VF₂/F₄ random copolymers. *J. Appl. Phys.* **73**, 1395–1402 (1993).

Acknowledgements. We thank P. A. Dowben and J. F. Scott for suggestions for improving the manuscript. Work at the University of Nebraska was supported by the USA NSF Division of Electronic and Communications Systems and by the Nebraska Research Initiative through the Center for Materials Research and Analysis. Work at the Institute of Crystallography was supported by INTAS.

Correspondence and requests for materials should be addressed to S.D. (e-mail: ducharme@unlinfo.unl.edu).

Surface-induced structure formation of polymer blends on patterned substrates

Martin Böltau, Stefan Walheim, Jürgen Mlyněk, Georg Krausch* & Ullrich Steiner

Fakultät für Physik, Universität Konstanz, 78457 Konstanz, Germany

Phase separation in bulk mixtures commonly leads to an isotropic, disordered morphology of the coexisting phases¹. The presence of a surface can significantly alter the phase-separation process, however^{2,3}. Here we show that the domains of a phase-separating mixture of polymers in a thin film can be guided into arbitrary structures by a surface with a prepatterned variation of surface energies. Such a pattern can be imposed on a surface by using printing methods for depositing microstructured molecular

films⁴, thereby allowing for such patterns to be readily transferred to a two-component polymer film. This approach might provide a simple means for fabricating polymer-based microelectronic circuits⁵ or polymer resists for lithographic semiconductor processing.

Our approach was quite general and works for a large number of binary polymer pairs and for a variety of different substrates. To identify an appropriate system, we studied the phase morphology of a polymer pair on homogeneous substrates of varying surface energies⁶. The experimental procedure was simple: the two polymers were dissolved in a common solvent, a drop of the solution was placed on a substrate, and the thin film was prepared by rapid rotation of the substrate (spin-coating). As the polymers are strongly incompatible, demixing takes place during the spin-coating procedure. The domain morphology that forms during solvent evaporation was observed using atomic force microscopy (AFM) (Fig. 1). Thin films of a polystyrene (PS)–polyvinylpyridine (PVP) mixture were spun-cast onto a gold (Au) substrate which was partially covered by a self-assembled monolayer (SAM). As PS and PVP are strongly incompatible, coexisting phases of the almost pure polymer components are found after film preparation⁶. The AFM image shows a homogeneous film surface on the bare Au substrate (top, Fig. 1a), whereas a corrugated film topography is found on the SAM-covered part of the substrate. The friction mode image (Fig. 1b), which provides a material contrast at the film surface⁷, reveals that these surface undulations coincide with a lateral composition variation at the surface of the polymer film. The Au surface (top, Fig. 1b), which shows no contrast in the friction image, is covered by only one polymer species. For infor-

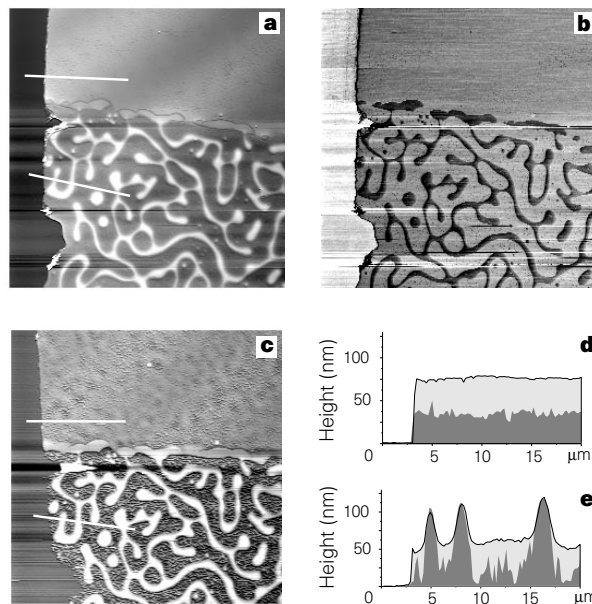


Figure 1 AFM images ($50 \times 50 \mu\text{m}^2$) of a PS/PVP blend (50%:50% w/w) spun-cast from a tetrahydrofuran (THF) solution (1.5% polymer by weight) onto a Au surface (50 nm of Au evaporated onto a silicon oxide surface). **a**, Topography image of the sample as cast. The friction mode image⁷ in **b** exhibits a material contrast between PS and PVP at the surface. In **c**, the PS phase was removed by dissolution in cyclohexane and an AFM topography image of the remaining PVP was taken at the same location, as in **a**. The images show a different domain morphology for the top half (bare Au surface) and the bottom half of the sample where the Au surface was hydrophobized by depositing a self-assembled monolayer (SAM) of octadecylmercaptan before polymer deposition⁴. In **d**, **e**, the PS (light grey) and PVP (dark grey) phase distribution in the film is visualized by a superposition of cross-sections at the locations indicated by lines in **a** and **c**. The bilayer in **d** forms on the more polar gold surface, whereas a laterally separated phase morphology is found on the SAM-covered substrate.

* Present address: Institut für Physikalische Chemie, Ludwig Maximilians Universität München, Theresienstrasse 39, 80333 München, Germany.

mation on the polymer composition inside the film, the PS phase can be removed by immersing the sample in cyclohexane, which is a solvent for PS, but does not dissolve PVP. In Fig. 1c, the AFM image shows the remaining PVP phase after removal of the PS. This image exhibits the same lateral structure as that in Fig. 1a, with an increased vertical contrast. The film composition normal to the substrate surface is displayed as a superposition of cross-sections from Fig. 1a and c. For the homogeneous film areas, a PS/PVP bilayer is formed, with the more polar PVP next to the Au substrate surface. On the nonpolar SAM surface, no preferential adsorption takes place, and a laterally disordered domain morphology of PS and PVP is seen (Fig. 1e). The strong topographic variation of the

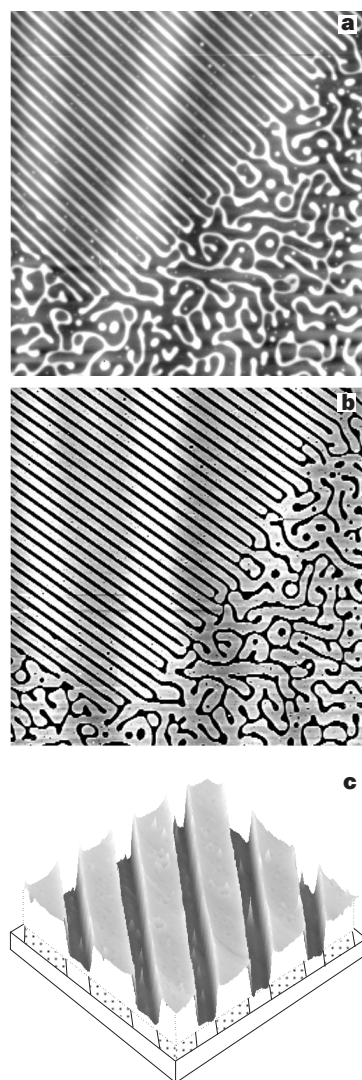


Figure 2 The same PS/PVP blend as in Fig. 1, spun-cast on a patterned Au substrate. The lateral surface pattern was created by microcontact printing, whereby a polydimethyl-siloxane stamp is soaked in an octadecyl mercaptan solution and placed onto a Au surface. A self-assembled monolayer forms only on surface regions where the stamp touches the substrate surface. The resulting Au/SAM pattern ($2.4\text{ }\mu\text{m}$ periodicity) features stripes of alternating surface energy with only little topographic variation. The AFM images show a $80 \times 80\text{ }\mu\text{m}^2$ area: **a**, as cast; **b**, after removal of PVP by dissolution in ethanol. The PS and PVP phases are organized in a surface-induced fashion in which PVP lies on the bare Au surface regions, displacing the PS onto the SAM stripes. The perspective representation in **c** visualizes the rectangular cross-section of the 65-nm-high PS stripes in **b**. The dotted areas in **c** indicate the SAM-covered substrate regions. In the bottom parts of **a** and **b**, where the substrate is completely SAM covered, the disordered structure is the same as that in Fig. 1.

domain morphology on the SAM-covered part of the surface (Fig. 1a, e) is due to the different solubility of the two polymers in their common solvent⁶.

The results shown in Fig. 1 suggest a strategy for the transfer of a lateral variation in surface energy into a composition pattern of a polymer film. To create a surface with laterally varying surface energies, micro-contact printing technology was used⁴. After spin-coating a PS/PVP solution onto the prepatterned surface, the image in Fig. 2a is observed. The surface corrugation from Fig. 1a is now perfectly aligned in a grid pattern. As in Fig. 1, the topographic undulation originates from a lateral organization of the polymer phases, as confirmed by friction-mode AFM (results not shown). This is corroborated by removing the PVP phase by immersing the sample in ethanol, which is a selective solvent for PVP. The remaining PS domains show a remarkable contrast in the AFM image (Fig. 2b): stripes with a rectangular cross-section. The polymer domain morphology in Fig. 2 is due to the strong affinity of the PVP to the more polar Au surface, displacing the PS towards the SAM-covered parts of the substrate. As in Fig. 1a and e, the PVP stripes protrude from the film surface, giving rise to the topographic image observed in Fig. 2a. This experiment demonstrates a transfer of a difference in surface energies into a composition variation of a binary polymer film. The perfect replication process extends over much larger length scales than those shown in the AFM images and is limited only by the sample size. Note that the rapid quench by solvent evaporation leads to domain structures far from their

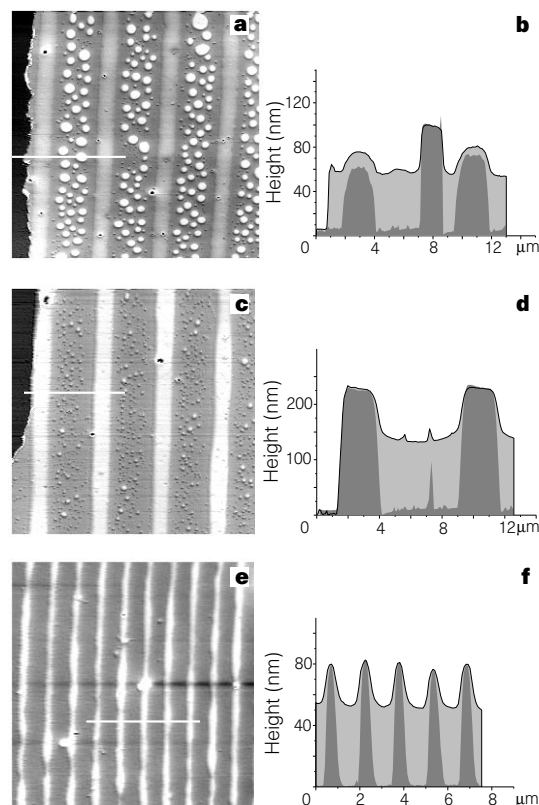


Figure 3 Films of PS and PSBr (50:50 w/w) spun-cast from toluene solution (2–4% by weight) onto a patterned silicon wafer which features alternating stripes of oxidized (SiO_x) and hydrogen-terminated silicon (SiH). The substrate pattern was generated by standard optical lithography combined with an etch in hydrofluoric acid. AFM images of the film as cast (**a**, **c**) were $30 \times 30\text{ }\mu\text{m}^2$, and $15 \times 15\text{ }\mu\text{m}^2$ (**e**). In **b**, **d** and **f**, cross-sections are shown from the locations indicated by lines in **a**, **c** and **d**, and corresponding images where the PS was removed by dissolution in cyclohexane: PS, light grey; PSBr, dark grey. The imperfect order in **a** can be improved by either changing the film thickness from 65 to 175 nm (**c**) or by reducing the periodicity of the SiO_x/SiH grid from 8 to $1.5\text{ }\mu\text{m}$ (**e**).

Table 1 Molecular characteristics of polymers

Polymer	M_w	Polydispersity index
PS	94.9 K	1.06
PVP	115 K	1.03
PSBr	215 K	1.02

M_w , weight average molecular weight. All materials were used as obtained from the Polymer Standards Service (Mainz, Germany). The degree of bromination of the PSBr was 52.2 mol %.

thermodynamic equilibrium. The rectangular cross-section of the PS phase in Fig. 2c is a direct consequence of this rapid solidification of the binary polymer blend⁶. In contrast, wetting experiments of a single liquid on a patterned substrate show a more rounded topography⁸.

These results were achieved only after careful optimization of several experimental parameters. For a given pattern periodicity, perfect lateral organization of the polymer phases is observed only for a given film thickness. To explore the optimum conditions for the self-organization process and to demonstrate the universality of our approach, a second binary polymer mixture on a different prestructured substrate surface was investigated. Mixtures of PS and a partially brominated polystyrene (PSBr) were spun-cast onto patterned silicon oxide (SiO_x)/hydrogen-terminated silicon (SiH) surfaces with varying grating periodicities. AFM cross-sections of the as-cast films and of the remaining PSBr phases after removing the PS (data not shown) are superimposed to yield the cross-sectional domain morphologies. In Fig. 3a, b a partial surface directed reorganization of PSBr and PS domains has taken place, but the lateral organization is incomplete. Figure 3a, b reveals a phase behaviour that is very similar to the separation of a PS/PSBr mixture on homogeneous SiO_x and SiH surfaces, respectively: PS/PSBr bilayers on the SiH surface regions and a lateral PS/PSBr phase morphology with $\sim 1\text{-}\mu\text{m}$ -wide PSBr domains surrounded by the PS phase on SiO_x . As the film thickness is increased, the quality of the pattern replication successively improves (Fig. 3c, d). For a given film thickness (that is, 65 nm in Fig. 3a), on the other hand, a reduction of the substrate periodicity leads to a better lateral order of the domain morphology (Fig. 3e, f). The results shown in Fig. 3 suggest an underlying principle that governs the ordering process. We compared the $\sim 1\text{-}\mu\text{m}$ average size of the PSBr domains on the SiO_x surface regions in Fig. 3a with the 750-nm stripe width of the SiH/ SiO_x substrate pattern in Fig. 3e: this indicated that complete ordering of the polymer phases can occur only for pattern sizes comparable to the characteristic length of the lateral domain morphology on the respective homogeneous substrate. Although the details of the surface-induced demixing process are not yet known, the existence of a well defined lateral length scale suggests a bulk- or surface spinodal instability which develops during the spin-coating process. A quantitative study of phase separation during spin-coating is currently underway.

We have made use of two fundamental thermodynamic principles—phase separation of a binary fluid and selective adsorption from a binary mixture—to transpose a variation in surface energies of a flat substrate into a concentration variation of a binary polymer film. The micrometre size structures perfectly replicate the substrate pattern over lateral dimensions which are limited only by the sample size (1 cm^2 in our case). The size of the structures that can be produced is at present dictated by the availability of substrate templates, and lateral structure sizes of less than 100 nm should be feasible. We emphasize that the structures reported here are replicated from a prepatterned substrate, rather than driven by a molecular self-organization process, as for example in microphase separated block-copolymer melts⁹. In these systems, the morphology of the lateral structure is determined by the architecture of the constituent molecules, and long-range order is difficult to achieve¹⁰. But with our approach, a given experimental system (that is, a particular polymer blend) can be used to produce structures of

widely varying morphologies and length scales. In contrast to bare patterned SAM substrates, the patterned polymer films are thick enough to allow further semiconductor processing steps such as reactive ion etching. Moreover, the polymer layers can be lifted off the substrates, and several structured films can be superimposed. Therefore, patterned SAM surfaces can be used as templates for multiple pattern replication. We also note that the manufacture of the structured polymer films takes just a few minutes, is inexpensive, and, once optimized, is highly reproducible. From a technological viewpoint, various applications can be envisaged, such as the formation of photonic bandgap materials¹¹, patterned polymer surfaces with selective molecular recognition capabilities¹², and optical devices¹³. In addition to their technological relevance, lateral nanoscale structures in polymer films could provide a new tool to study phase equilibria near surfaces on a molecular level. □

Received 19 September; accepted 9 December 1997.

- Guntton, J. D., San Miguel, M. & Sahni, P. S. in *Phase Transitions and Critical Phenomena* (eds Domb, C. & Lebovitz, J. L.) Vol. 8, 267–466 (Academic, London, 1983).
- Jones, R. A. L., Norton, L. J., Kramer, E. J., Bates, F. S. & Wiltzius, P. Surface-directed spinodal decomposition. *Phys. Rev. Lett.* **66**, 1326–1329 (1991).
- Krausch, G., Kramer, E. J., Rafailovich, M. H. & Sokolov, J. Self assembly of a homopolymer mixture via phase separation. *Appl. Phys. Lett.* **64**, 2655–2657 (1994).
- Xia, Y., Zaho, X.-M. & Whitesides, G. M. Pattern transfer: Self-assembled monolayers as ultrathin resists. *Microelectr. Eng.* **32**, 255–268 (1996).
- Service, R. F. Patterning electronics on the cheap. *Science* **278**, 383–384 (1997).
- Walheim, S., Böltz, M., Mlynek, J., Krausch, G. & Steiner, U. Structure formation via polymer demixing in spin-cast films. *Macromolecules* **30**, 4995–5003 (1997).
- Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Abbot, N. L., Folkers, J. P. & Whitesides, G. M. Manipulation of the wettability of surfaces on the 0.1- to 1-micrometer scale through micromachining and molecular self assembly. *Science* **257**, 1380–1382 (1992).
- Morkved, T. L., Wiltzius, P., Jaeger, H. M., Grier, D. G. & Witten, T. A. Mesoscopic self-assembly of gold islands on diblock-copolymer films. *Appl. Phys. Lett.* **64**, 422–424 (1994).
- Morkved, T. L. et al. Local control of microdomain orientation in diblock copolymer thin films with electric fields. *Science* **273**, 931–933 (1996).
- Haus, J. W. in *Quantum Optics of Confined Systems* (eds Ducloy, M. & Block, D.) 101–141 (Kluwer Academic, Dordrecht, 1996).
- Prime, K. L. & Whitesides, G. M. Self-assembled organic monolayers: Model systems for studying adsorption of proteins at surfaces. *Science* **252**, 1164–1167 (1991).
- Halls, J. et al. Efficient photodiodes from interpenetrating networks. *Nature* **376**, 498–500 (1995).

Acknowledgements. This work was supported by the Deutsche Forschungsgesellschaft (DFG), the Volkswagenstiftung, and NATO. U.S. acknowledges the financial support from a research fellowship (Habilitation-Stipendium) of the DFG. We thank E. J. Kramer and J. Heier for their help preparing the PDMS stamps.

Correspondence and requests for materials should be addressed to U.S. (e-mail address: Ulli.Steiner@uni-konstanz.de).

Origin of upper-ocean warming and El Niño change on decadal scales in the tropical Pacific Ocean

Rong-Hua Zhang*, Lewis M. Rothstein* & Antonio J. Busalacchi†

* Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

† Laboratory for Hydrospheric Process, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

The cause of decadal-scale variability in the tropical Pacific Ocean—such as that marked by the 1976–77 shift in the El Niño/Southern Oscillation^{1–7}—is poorly understood. Unravelling the mechanism of the recent decade-long warming in the tropical upper ocean is a particularly important challenge, given the link to El Niño variability, but establishing the hypothesized inter-annual/decadal oceanic connections between middle latitudes and tropics has proved elusive⁸. Here we present observational evidence that Pacific upper-ocean warming and decadal changes in the El Niño/Southern Oscillation after 1976 may originate from

decadal mid-latitude variability. In the middle 1970s the North Pacific Ocean is observed to have undergone a clear phase-transition; a 'see-saw' subsurface temperature anomaly pattern that rotates clockwise around the subtropical gyre. At middle latitudes a subsurface warm anomaly formed in the early 1970s from subducted surface-waters and penetrated through the sub-

tropics and into the tropics, thus perturbing the tropical thermocline and driving the formation of a warm surface-water anomaly that may have influenced El Niño in the 1980s. The identification of this teleconnection of extratropical thermal anomalies to the tropics, through a subsurface ocean 'bridge', may enable improved prediction of decadal-scale climate variability.

We have examined recently available observations^{9–12} of the Pacific basin, looking for interrelationships between the extratropics and tropics on interannual and interdecadal scales. Atmospheric variables were obtained from a new analysis of surface marine anomalies derived from the Comprehensive Ocean–Atmosphere Data Sets (COADS) from January 1945 to December 1989¹³. The oceanic data were yearly *in situ* temperature anomalies at 13 standard levels from the sea surface to 400 m depth analysed by Levitus *et al.*⁹ The yearly anomaly fields were produced by subtracting

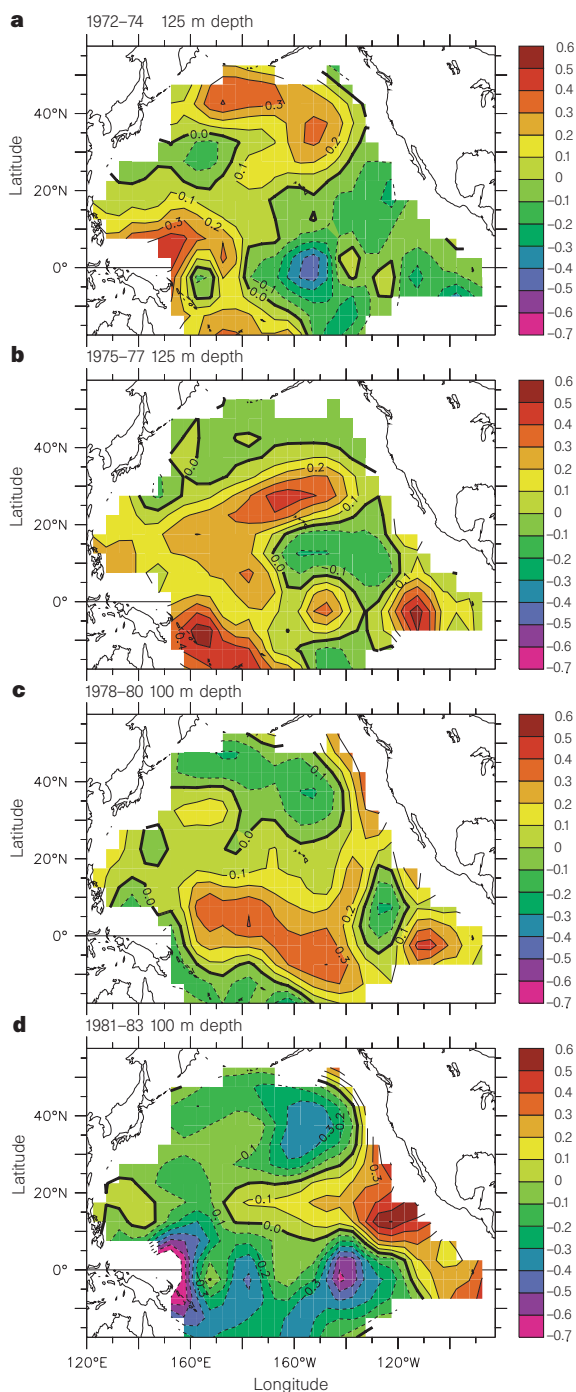


Figure 1 Horizontal distribution of ocean temperature anomalies averaged for four consecutive 3-year periods. Temperatures were measured at either 125 m depth (**a**, 1972–74; **b**, 1975–77) or at 100 m depth (**c**, 1978–80; **d**, 1981–83). The choice of different depths for detecting the basin-scale anomaly patterns of interest reflects the fact that the thermocline in the tropics is shallower than at middle latitudes. We note that some subsurface warm anomalies have penetrated from the middle latitudes into the tropics in **a** and **b**, and can be seen to extend eastwards in the tropics in **c** and **d**. The contour interval is 0.1 °C, with the colour bar on the right indicating anomalies from –0.7 to 0.6 °C.

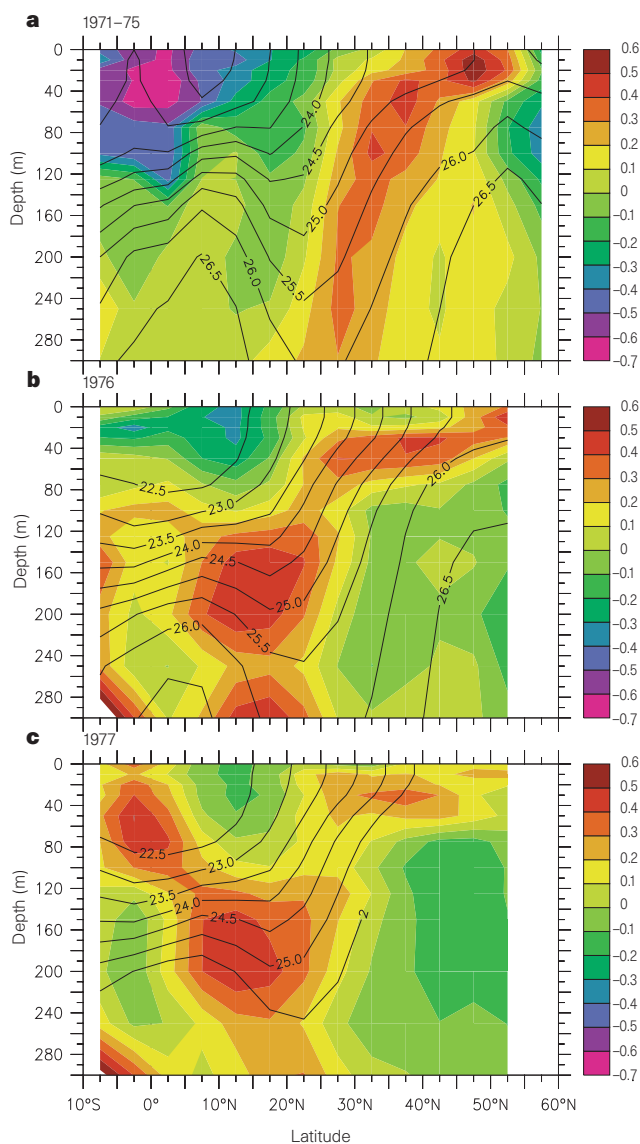


Figure 2 Meridional-depth sections of upper-ocean temperature anomalies (colour) and isopycnal surfaces (kg m^{-3} ; contour lines). Sections are along 157.5°W for years 1971–75 (**a**), and along 177.5°E for years 1976 (**b**) and 1977 (**c**). Two different longitudinal sections have been used to show clearly the signals of interest (for example, the warm anomaly subduction occurs mainly in the eastern North Pacific whereas the warm anomaly penetration from the subtropics into tropics occurs in the central basin). The colour bar on the right indicates temperature anomalies from –0.7 to 0.6 °C with intervals of 0.1 °C.

the averaged values from the 1961–90 mean values for the ocean, and from the 1961–89 mean values for the atmosphere.

To examine the spatial structure of decadal variability over the entire Pacific, we put together atmospheric and oceanic anomaly fields for three periods: from 1965 to 1972, from 1975 to 1977, and from 1980 to 1988. The choice of these averaged time periods is based on previous empirical orthogonal function (EOF) analyses of upper-ocean temperature anomalies as best representing the decadal warm and cold phases in the North Pacific, and their transition during 1975–77¹². During the period 1965–72, that is, the decadal warm phase (relative to the central middle latitudes), the Aleutian Low weakens, accompanied by anomalous anticyclonic circulation over the North Pacific, with easterly wind anomalies prevailing in the western tropical basin. Sea surface temperature (SST) is anomalously warm over the eastern and central middle latitudes, but is cold off the west coast of North America and in the tropical and subtropical regions. An abrupt shift of SST and atmospheric circulation occurs late in 1976^{1–6}, characterized by, among other things, a warming through much of the tropical central and eastern Pacific Ocean.

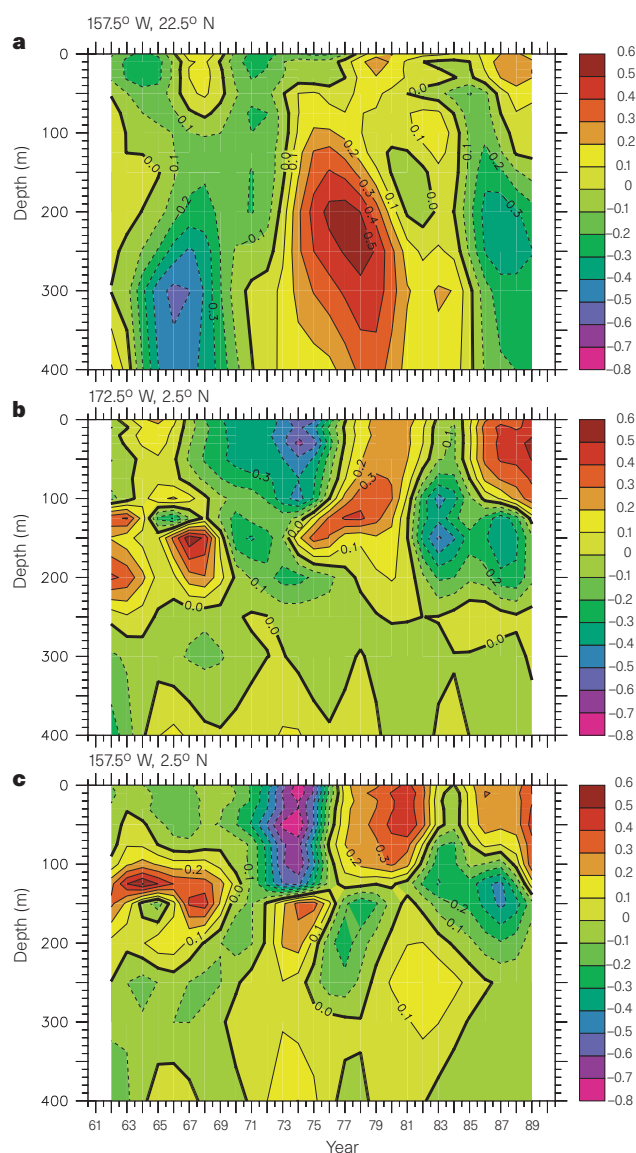


Figure 3 Upper ocean temperature anomalies as a function of depth and time in different locations. **a**, At 157.5° W and 22.5° N; **b**, at 172.5° W and 2.5° N; **c**, at 157.5° W and 2.5° N. A three-year running-mean filter has been applied to these data. The contour interval is 0.1 °C with the colour bar on the right indicating anomalies from –0.8 to 0.6 °C.

The upper-ocean temperature data set⁹ enables us to investigate the oceanographic component of this decadal variability over the entire Pacific; its basin-scale structure and evolution has been recently documented^{9,11,12}. Figure 1 shows the time-averaged temperature anomalies at subsurface depths in the sequence of four time periods. There is a well-defined pattern and an orderly space-time evolution around the subtropical gyre, with prominent planetary wave signatures in the subtropics^{12,14–16}. These features have been more clearly detected from EOF and regression pattern analyses¹². During the pre-1976–77 period, for about a decade, a large subsurface warm temperature anomaly covers the central middle latitudes, with cold anomalies in the subtropics (Fig. 1a). During the period 1976–77 (Fig. 1b), an apparent reversal of anomaly polarity begins to occur: the warm mid-latitude anomaly circulates southwestwards in the eastern subtropics while at the same time penetrating into the western subtropics and tropics, where there is also a corresponding warming. These subsurface anomalies continue to evolve on the basin scale (Fig. 1c, d).

The observed subsurface thermal anomalies tend to persist as they move around the basin and can be transported to the sea surface within strong upwelling and/or vertical mixing regions such as in the equatorial Pacific. The subsurface warm anomalies observed to have subducted in the middle latitudes in the late 1960s and early 1970s may therefore serve as the origin of upper-ocean warming (and decadal climate change) in the tropical Pacific in the 1980s.

In Figs 2–4 we show the chronological evidence for connections between surface and subsurface temperature anomalies at middle latitudes and in the tropics. Mid-latitude air–sea interactions produce persistent warm anomalies at the sea surface that are subducted into the thermocline (Fig. 2a) and then move westward and equatorward around the subtropical gyre on isopycnal surfaces^{10,17}. The subducted North Pacific thermocline warm anomaly, observed first at middle latitudes in the late 1960s and early 1970s, starts to circulate toward the Equator at subsurface ocean depths. In the middle 1970s, a strong warm anomaly has detached from the surface and is found in the central subtropics at depth (Fig. 3a).

Subsurface midlatitude-to-tropical ocean connections are further demonstrated in Fig. 4. During the 1960s and until about 1974, temperature variations are out of phase in the middle latitudes and tropics: a warm anomaly in the north, and a mainly cold anomaly to the south (Fig. 4a, b). Such out-of-phase variations are also evident during the 1980s, but with opposite sign. Throughout the 1970s and early 1980s, a slow southward propagation of a warm anomaly can be seen from the middle latitudes through the subtropics and into tropics (Fig. 4c). By the early 1970s, a related subsurface warm anomaly has already occurred in the far western tropical Pacific (Fig. 1a). However, a clear incursion of the extratropical warm anomaly into the tropics takes place in the middle 1970s (Fig. 2b, c).

We show the links between tropical subsurface and surface temperature anomalies in Fig. 2b, c and Fig. 3b, c, respectively. In 1976 (Fig. 2b), the largest warm anomaly (+0.4 °C) is located at 140–220 m depth between 8° N and 23° N in the central Pacific, with cold surface anomalies in the tropics. For the year 1977, the three-dimensional structure of temperature anomalies in a depth–latitude section can be seen in Fig. 2c. A significant warm anomaly is centred at ~180 m depth in the off-equatorial region (at ~13° N). Having become detached from the chief off-equatorial warm-anomaly centre, another warm anomaly can be seen to form in the equatorial region at relatively shallow depths (at ~70 m) in the central Pacific. The time–depth evolution is shown in Fig. 3b. A tail structure is clearly evident, with a warm anomaly originating from subsurface depths and spreading upwards with time. Based on our more complete and detailed analyses (not shown), the shallow equatorial warm anomaly is most likely to have come from the North Pacific subtropics, with anomalies at the sea surface lagging those at depth and to the west on the Equator. This anomaly seems

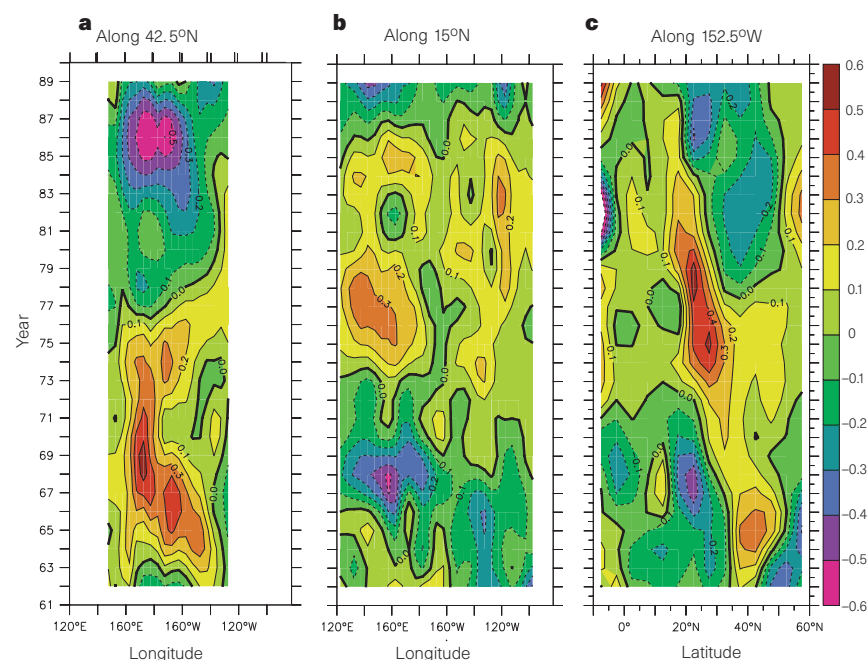


Figure 4 Ocean temperature variations with time from 1960 to 1990 at 250m depth. The data are plotted against longitude along 42.5°N (**a**), against longitude along 15°N (**b**), and against latitude along 152.5°W (**c**), demonstrating subsurface connections between temperature anomalies at middle latitudes and in the tropics. The contour interval is 0.1 °C, with dashed lines for cold anomalies. A three-year running mean filter has been applied to these data.

to be responsible for changes in the background thermal state and the behaviour of El Niño/Southern Oscillation (ENSO) immediately after 1976. Indeed, on the Equator, clear phase relationships between decadal anomalies of subsurface temperature and SST have been reported¹¹, with subsurface warm anomalies in the west preceding large-scale surface warming to the east. Further data analyses (not shown) indicate that decadal subsurface temperature variability in the tropics is not simply forced by tropical surface wind anomalies, which, in fact, are largely a response to SST changes. In addition, the subsurface warm anomalies appear to extend eastward along the Equator (Fig. 1b, c and Fig. 3b, c) and continue to move northward along the eastern boundary and, subsequently, westward away from the boundary in the subtropics as observed in the early 1980s (Fig. 1d).

So there are links between extratropical and tropical temperature anomalies through a subsurface ocean 'bridge' for the middle and late 1970s. The decadal warm-anomaly incursion from mid-latitude subduction sites into the tropics seems to form a warm thermocline anomaly in the western and central tropics. Once the tropical subsurface thermal structure has been perturbed, modelling studies have demonstrated how this temperature anomaly at depth can be communicated to the surface to influence SST in regions of strong vertical mixing and upwelling^{18,19}. Thus the extratropical warm anomaly may induce upper-ocean warming in the central and eastern tropics, which, in turn, can produce surface westerly wind anomalies to the west, setting up air-sea interactions. This helps to explain the observed decadal-scale ENSO variability in the 1980s, that is, the extremely strong 1982–83 ENSO, and the frequent occurrence of the El Niño warm phase but a lack of a cold phase.

These findings reported here have significant implications from a variety of perspectives. Many low-frequency climate signals have been ascribed to El Niño in the tropics, while the role of the mid-latitude ocean has been less studied. In particular, the 1976–77 Pacific climate shift in the middle and high latitudes has been explained as a decadal effect of tropical interannual variations^{1,4–6}. In fact, variations in the tropical Pacific are dominated by inter-annual signals²⁰; there are no potential mechanisms or processes which have been identified that could maintain a self-sustained tropical decadal oscillation. This suggests that observed decadal tropical variability may be better traced to decadal extratropical processes. As demonstrated here, before the tropical warming in the middle and late 1970s, a decadal transition from a warm to cold

phase occurs first in the North Pacific. Thus, thermocline anomalies associated with mid-latitude subduction processes may play an important role in perturbing tropical low frequency variability.

Our work also provides a basinwide, observational perspective on Pacific climate variability which suggests that temperature anomalies associated with the mid-latitude thermocline variability can penetrate all the way into the tropics. This is consistent with recent model studies of water pathways in the subtropical-tropical Pacific Ocean^{21–23}. Identification of such anomaly patterns in the ocean, as well as in the atmosphere²⁴, is important for understanding and prediction of decadal climate variability. For example, recent modelling studies of ENSO predictability^{25,26} have identified a decadal scale variability in prediction skill. It has been suggested that such changes may arise if there were changes to the background state of the tropical Pacific Ocean. Here we have seen a link between the middle latitudes and tropics that offers the potential of using the lead time intrinsic to the North Pacific subsurface thermal pattern to forecast decadal ENSO variability in the tropics.

The unusual warming in the tropical Pacific Ocean observed in the early 1990s^{27,28} appears to be different from the warmings of the 1970s and 80s. It seems likely, for instance, that the propagation of cold extratropical anomalies (Fig. 4c) into the tropics in the late 1980s would introduce a cooling tendency in the tropical Pacific Ocean. The nature of the unexpected warming in the early 1990s has indeed been a topic of much debate^{27,28}; is it a forced global warming signature or is it natural climate variability? Different decadal variability modes and the effects of the South Pacific Ocean are also possible. Sufficient temporal data coverage is not yet available to resolve these questions. □

Received 24 February; accepted 15 December 1997.

1. Trenberth, K. E. & Hurrell, J. W. Decadal atmosphere-ocean variations in the Pacific. *Clim. Dyn.* **9**, 303–319 (1994).
2. Ebbsmeyer, C. C. *et al.* in *Proc. 7th Ann. Pacific Climate Workshop* (Interagency Ecol. Stud. Prog. Rep. 26, Calif. Dept. of Water Resource, Los Angeles, 1991).
3. Nitta, T. & Yamada, S. Recent warming of tropical sea surface temperature and its relationship to the Northern Hemisphere circulation. *J. Meteorol. Soc. Jpn* **67**, 375–383 (1989).
4. Graham, N. E. Decadal scale variability in the tropical and North Pacific during the 1970s and 1980s: observations and model results. *Clim. Dyn.* **10**, 135–162 (1994).
5. Miller, A. J. *et al.* The 1976–77 climate shift of the Pacific Ocean. *Oceanography* **7**, 21–26 (1994).
6. Miller, A. J. *et al.* Interdecadal variability of the Pacific Ocean: model response to observed heat flux and wind stress anomalies. *Clim. Dyn.* **9**, 287–302 (1994).
7. Wang, B. Interdecadal changes in El Niño onset in the last four decades. *J. Clim.* **8**, 267–285 (1995).
8. Kerr, R. A. Unmasking a shifty climate system. *Science* **255**, 1508–1510 (1992).
9. Levitus, S., Boyer, T. P. & Antonov, J. *World Ocean Atlas Vol. 5, Interannual Variability of Upper Ocean Thermal Structure* (Atlas NESDIS 5, NOAA, Silver Spring, 1994).

10. Deser, C., Alexander, M. A. & Timlin, M. S. Upper ocean thermal variations in the North Pacific during 1970–1991. *J. Clim.* **9**, 1840–1855 (1996).
11. Zhang, R.-H. & Levitus, S. Structure and evolution of interannual variability of the tropical Pacific upper ocean temperature. *J. Geophys. Res.* **101**, 20501–20524 (1996).
12. Zhang, R.-H. & Levitus, S. Structure and cycle of decadal variability of upper ocean temperature in the North Pacific. *J. Clim.* **10**, 710–727 (1997).
13. Da Silva, M., Young, C. C. & Levitus, S. *Atlas of Surface Marine Anomalies* Vol. 2, *Directly Observed Quantities* (Atlas NESDIS 7, NOAA, Silver Spring, 1994).
14. Jacobs, G. A. *et al.* Decade-scale trans-Pacific propagation and warming effects of an El Niño anomaly. *Nature* **370**, 360–363 (1994).
15. Latif, M. & Barnett, T. P. Causes of decadal climate variability over the North Pacific and North America. *Science* **266**, 634–637 (1994).
16. Latif, M. & Barnett, T. P. Decadal climate variability over the North Pacific and North America: dynamics and predictability. *J. Clim.* **9**, 2407–2423 (1996).
17. Gu, D.-F. & Philander, S. G. H. Interdecadal climate fluctuations that depend on exchanges between the tropical and extratropics. *Science* **275**, 805–807 (1997).
18. Hirst, A. C. & Godfrey, J. S. The role of Indonesia throughflow in a global ocean GCM. *J. Phys. Oceanogr.* **23**, 1057–1086 (1993).
19. Liu, Z. & Philander, S. G. H. How different wind stress affect the tropical-subtropical circulations of the upper ocean. *J. Phys. Oceanogr.* **25**, 449–462 (1995).
20. Philander, S. G. H. *El Niño, La Niña and the Southern Oscillation* (Academic, New York, 1990).
21. Liu, Z. A simple model of the mass exchange between the subtropical and tropical ocean. *J. Phys. Oceanogr.* **24**, 1153–1165 (1994).
22. McCreary, J. P. & Lu, P. Interaction between the subtropical and equatorial ocean circulations: the subtropical cell. *J. Phys. Oceanogr.* **24**, 466–497 (1994).
23. Rothstein, L. M., Zhang, R.-H., Busalacchi, A. J. & Chen, D. A numerical simulation of the mean water pathways in the subtropical and tropical Pacific Ocean. *J. Phys. Oceanogr.* **28**, 322–342 (1998).
24. Bjerknes, J. Atmospheric teleconnections from the equatorial Pacific. *Mon. Weath. Rev.* **97**, 163–172 (1969).
25. Balmaseda, M. A., Davey, M. K. & Anderson, D. L. T. Decadal and seasonal dependence of ENSO prediction skill. *J. Clim.* **8**, 2705–2715 (1995).
26. Chen, D., Zebiak, S. E., Busalacchi, A. J. & Cane, M. A. An improved procedure for El Niño forecasting: Implications for predictability. *Science* **269**, 1699–1702 (1995).
27. Trenberth, K. & Hoar, T. The 1990–95 El Niño–South Oscillation event: Longest on record. *Geophys. Res. Lett.* **23**, 57–60 (1996).
28. Latif, M., Kleeman, R. & Eckert, C. Greenhouse warming, decadal variability, or El Niño? an attempt to understand the anomalous 1990s. *J. Clim.* **10**, 2221–2239 (1997).

Acknowledgements. We thank S. Levitus and other staff at the Ocean Climate Laboratory (NODC/NOAA) for their help and support in analysing observational data, and also thank M. Latif, Z. Liu, S. Tett, C. Deser, X.-Z. Liang and J. Picaut for comments. This work was supported by the NSF, NOAA and NASA.

Correspondence and requests for materials should be addressed to R.-H.Z. (e-mail address: zrh@sequan.gso.uri.edu).

Evidence for pressure-release melting beneath magmatic arcs from basalt at Galunggung, Indonesia

T. W. Sisson* & S. Bronto†

* US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

† Volcanological Survey of Indonesia, Jl. Diponegoro no. 57, Bandung, 40122 Indonesia

The melting of peridotite in the mantle wedge above subduction zones is generally believed to involve hydrous fluids derived from the subducting slab¹. But if mantle peridotite is upwelling within the wedge, melting due to pressure release could also contribute to magma production. Here we present measurements of the volatile content of primitive magmas from Galunggung volcano in the Indonesian arc which indicate that these magmas were derived from the pressure-release melting of hot mantle peridotite. The samples that we have analysed consist of mafic glass inclusions in high-magnesium basalts. The inclusions contain uniformly low H₂O concentrations (0.21–0.38 wt%), yet relatively high levels of CO₂ (up to 750 p.p.m.) indicating that the low H₂O concentrations are primary and not due to degassing of the magma. Results from previous anhydrous melting experiments on a chemically similar Aleutian basalts² indicate that the Galunggung high-magnesium basalts were last in equilibrium with peridotite at ~1,320 °C and 1.2 GPa. These high temperatures at shallow sub-crustal levels (about 300–600 °C hotter than predicted by geodynamic models^{1,3}), combined with the production of nearly H₂O-free

basaltic melts, provide strong evidence that pressure-release melting due to upwelling in the sub-arc mantle has taken place. Regional low-potassium⁴ and low-H₂O (ref. 5) basalts found in the Cascade arc indicate that such upwelling-induced melting can be widespread.

Galunggung, an arc magmatic-front volcano in western Java, Indonesia, erupted in 1982–83 (ref. 6). Activity began with andesitic pyroclastic flows and products became more mafic with time, ending in several months of strombolian and effusive eruptions of high-Mg basalt (10.0–11.8 wt% MgO). These basalts have high concentrations of compatible trace elements⁷, have compositions similar to Aleutian arc high-Mg basalts (Table 1), and carry phenocrysts of olivine to Fo₉₁ (although most are Fo₈₈; Fo = 100 Mg/(Mg + Fe)), as well as diopside, Cr-spinel and plagioclase. The primitive compositions of the high-Mg basalts, their primitive phenocrysts, and their mantle-like radiogenic and stable-isotope values indicate that they are near-primary magmas supplied to the volcanic front in western Java^{7,8}.

Inclusions of mafic glass are common in olivine phenocrysts of the high-Mg basalts, and are trapped samples of pre-eruptive melts. Other included phases are Cr-spinel and Al-enstatite (which is rare). These glass inclusions are generally smaller than 50 µm in diameter, and each contains a single vapour bubble decorated with tiny Fe-sulphide blebs; most glass inclusions also contain a single grain of Cr-spinel. Olivine phenocrysts that host glass inclusions are weakly zoned in their interiors and have compositions in the range Fo_{89–73}, although most are >Fo₈₃ and have NiO contents of 0.31–0.10 wt%. Glass inclusions from two high-Mg basaltic bombs were analysed for major-element, trace-element and volatile contents.

Inclusion glasses range continuously from transitional basaltic to strongly undersaturated compositions with low SiO₂ contents, high CaO contents, and high CaO/Al₂O₃ ratios, but with alkali contents similar to the transitional basaltic inclusions (Table 1). Glass inclusions have lower MgO contents and Mg/(Mg + Fe) than the host high-Mg basalts, due in part to 5–13 wt% post-entrapment crystallization of olivine on the inclusion walls. Projected from olivine (Fig. 1), to factor out the effects of host-olivine crystallization, the inclusion glasses form an array from near the Galunggung high-Mg basalts (basaltic inclusions) to the diopside–Ca-tschermak–jadeite compositional plane (strongly undersaturated

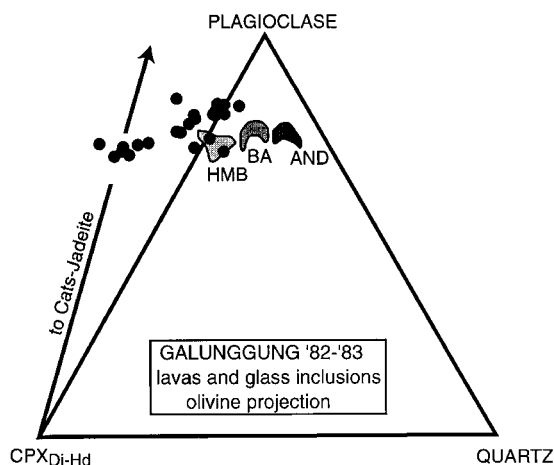


Figure 1 Galunggung 1982–83 eruption whole rocks (fields) (ref. 7 and this work) and high-Mg-basalt inclusion glasses (filled circles) projected²⁴ from olivine onto the plagioclase-clinopyroxene (cpx)–quartz compositional plane. Late-erupted high-Mg basalts (HMB), light shading; basaltic andesites (BA), medium shading; and early-erupted andesites (AND), dark shading. CPX_{Di-Hd} is diopside + hedenbergite; arrow shows trajectory to calcium-tschermak (cats) and jadeite pyroxene components. Whole-rock compositions were measured by X-ray fluorescence at the University of Canterbury, NZ. Glass and mineral inclusions and phenocrysts were analysed by electron-microprobe at the USGS, Menlo Park, CA.

inclusions). Similar projected positions of high-Mg basalts and basaltic inclusions, as well as similar ratios of incompatible components (average Mg-basalt $K_2O/Na_2O = 0.15$, $TiO_2/Na_2O = 0.37$; average basaltic inclusion $K_2O/Na_2O = 0.16$, $TiO_2/Na_2O = 0.39$), support a close genetic relation of basaltic inclusions and host basalts. Following correction for host-olivine crystallization, the high-Mg basalts and basaltic inclusions can be related (by mass balance) as parent and daughter by deep fractionation of diopsidic clinopyroxene and olivine followed by mixing with up to 25 wt% of strongly undersaturated melt. As with other forsteritic olivine-hosted melt inclusions, Galunggung melt inclusions are depleted in Fe either by diffusion into olivine^{9,10}, or by separation of an Fe-rich phase (perhaps Fe sulphide), or by assimilation of forsteritic olivine during deep fractionation. The strongly undersaturated inclusions resemble ankaramites in their high CaO contents, CaO/Al₂O₃ ratios, and low SiO₂ and alkali contents, although no ankaramites have erupted near Galunggung.

Compared to normal mid-ocean-ridge basalt (MORB), Galunggung high-Mg basalts have small depletions of high-field-strength elements relative to rare-earth elements, but have distinct enrichments of Sr, K and Ba (Fig. 2). Trace-element abundance patterns of Galunggung glass inclusions closely resemble those of the host high-Mg basalts, which further supports their genetic link. However, the diversity in inclusion major-element abundances precludes the possibility that included melts were liquids derived from crystallization of a single parent magma. The arc elemental signature of the glass inclusions is generally more pronounced than in most of the high-Mg basalts, some of which have Ba/Zr values close to the global MORB–ocean island basalt array (Fig. 2) and may have restricted contributions of subduction components.

Concentrations of H₂O in inclusion glasses are consistently low, from 0.21 to 0.38 wt% (Table 1, Fig. 2), regardless of inclusion type. CO₂ concentrations are highest (to 750 p.p.m.) in the most strongly undersaturated inclusions, but some basaltic inclusions also have

high CO₂ concentrations (to 360 p.p.m.). The high CO₂ contents of inclusion glasses show that their low H₂O contents are not due to degassing, which depletes melts in CO₂ first before lowering their H₂O contents appreciably¹¹. Diffusion through host olivine is also unlikely to have caused the uniform and low H₂O contents of melt inclusions, as olivine-hosted melt inclusions at some other arc volcanoes show strong correlations between inclusion H₂O contents and indices of differentiation and degassing,^{5,12} and marked H₂O diffusion would destroy such correlations. Because CO₂ is more soluble in undersaturated liquids than in tholeiitic liquids¹³, the higher CO₂ contents of the undersaturated inclusions may result from preferential retention of CO₂ during vesicle growth or enhanced dissolution of CO₂ during melting, as opposed to an elevated CO₂ content of their source. Inclusion S concentrations (undersaturated, 2,200–1,600 p.p.m.; basaltic, 1,600–500 p.p.m.) are higher to much higher than those of subaerially degassed basaltic melts (≤ 250 p.p.m.; ref. 14), which further indicates that the low H₂O contents of the inclusions are not artefacts of degassing. Chlorine concentrations are high (700–1,600 p.p.m., averaging 1,200 p.p.m.) in the glass inclusions and are not related to the major-element compositions of the inclusions. Similarly high Cl concentrations are common in other mafic arc melts^{5,15,16}, but the low H₂O concentrations at Galunggung lead to H₂O/Cl values (near 3) that are the lowest yet reported for undegassed mafic arc liquids.

Galunggung high-Mg basalts are close to primary liquid compositions, an interpretation based on apparent Fe–Mg equilibrium between the common forsteritic olivine phenocrysts and whole-rock compositions^{7,8}, as well as the similar olivine-projected positions of basaltic glass inclusions and host high-Mg basalts (Fig. 1). The low H₂O contents of the glass inclusions and the similarities between basaltic inclusions and host magmas strongly suggest that the high-Mg basalt melts also had low H₂O contents, similar to those of MORB liquids¹⁷. Galunggung high-Mg basalt melts were probably last in equilibrium with peridotite at $\sim 1,320^\circ\text{C}$ and

Table 1 Representative whole-rock and glass-inclusion compositions

	Rocks		Glass inclusions				
	ID-16	Avg-MGB	B1-1	G-2	J-1	F-1	I-E
SiO ₂ (%)	48.9	49.1 ± 0.2	50.0	51.0	45.0	43.9	44.2
TiO ₂	0.70	0.82 ± 0.1	1.27	1.21	1.20	1.28	1.30
Al ₂ O ₃	16.0	15.9 ± 0.3	20.8	20.4	19.2	18.4	19.2
FeO*	8.90	9.0 ± 0.1	5.32	5.48	6.13	7.64	8.73
MgO	11.4	10.5 ± 0.4	4.38	4.66	4.39	4.99	4.10
MnO	0.17	0.17 ± 0.01	0.11	0.10	0.09	0.09	0.10
CaO	10.89	11.3 ± 0.1	13.0	13.0	18.7	18.4	17.7
Na ₂ O	2.21	2.21 ± 0.07	3.34	3.52	3.07	3.49	2.70
K ₂ O	0.52	0.34 ± 0.01	0.60	0.58	0.53	0.66	0.42
P ₂ O ₅	0.12	0.10 ± 0.01	0.12	0.14	0.14	0.18	0.12
H ₂ O	–	–	0.30	0.38	0.30	0.38	i.f.
CO ₂ (p.p.m.)	–	–	<30	360	500	510	i.f.
S	–	–	920	620	1,930	2,240	1,690
Cl	–	–	1,050	1,100	1,600	950	1,190
total (%)	99.8	99.4	99.4	100.7	99.2	99.8	98.9
Ba (p.p.m.)		60 ± 18	140	137	153	295	121
Sr		253 ± 67	275	294	284	382	251
Y		17 ± 1	22.8	21.4	25.7	28.4	23.8
Zr		46 ± 7	62.6	62.1	64.7	74.0	53.8
Nb		2 ± 1	2.4	2.7	2.4	3.0	1.9
Sc		39 ± 1	40.9	34.1	53.8	81.6	52.1
V		272 ± 5	361	340	356	398	343
Cr		568 ± 58	17.0	–	24.9	15.3	21.2
La		3.9 ± 0.1	5.83	5.96	6.57	8.45	5.09
Ce		9.4 ± 1.5	14.2	14.5	15.6	19.8	12.9
Nd		–	10.0	10.1	12.3	14.3	10.6
Sm		1.9 ± 0.1	3.28	2.93	3.96	4.59	3.42
Eu		0.82 ± 0.03	0.81	1.00	1.11	1.16	1.01
Dy		–	4.15	4.13	4.72	5.31	4.13
Er		–	2.24	2.22	2.85	2.95	2.39
Yb		2.13 ± 0.17	2.33	2.20	2.53	2.84	2.49
Host Fo (%)			88.2	87.8	89.1	88.2	84.4

Samples: ID-16, Aleutian high-Mg basalt²²; Avg-MGB, average Galunggung high-Mg basalt (this study and ref. 7), ± gives 1σ variations; inclusions B, G, J, F and I were separated from high-Mg basalt sample Gal-20298; i.f., samples in which interference fringes prevented accurate H₂O and CO₂ analyses. Host forsterite contents were measured within 10 μm of the associated inclusions. Whole-rock Nb is derived from Ta (ref. 7) and arc Nb–Ta systematics²³. FeO* is all iron reported as FeO; numbers after inclusion letters are sample designations only.

1.2 GPa, as shown by anhydrous experiments on a chemically similar Aleutian high-Mg basalt² (Table 1). This inferred temperature is much higher than is predicted for that pressure by dynamic models of the sub-arc mantle in which peridotite flows towards the subducting slab at shallow levels and then descends parallel to the slab¹. A similar temperature disparity exists for the Cascade arc, where low-K basalts with very low pre-eruptive H₂O contents⁵ last equilibrated with mantle peridotite at ~1,290 °C and 1.1 GPa (ref. 18).

High temperatures at shallow sub-crustal levels and the production of nearly H₂O-free basaltic melts are explained most simply by upwelling and pressure-release melting of hot peridotite in the mantle wedge. Upwelling must be a regional feature of the Cascade arc because the low-K, low-H₂O basalts have erupted along the volcanic front for a distance of at least 640 km (ref. 4). Upwelling is probably also present beneath much of the Java arc, in so far as

Galunggung is a magmatic-front volcano that has erupted andesites like those of other Javanese volcanoes through most of its history, and in so far as Galunggung does not occupy any tectonically unusual position that would suggest anomalous magma generation processes. That mantle upwells beneath arcs at some times is apparent from arcs that have rifted apart to form remnant arcs and back-arc basins. Mantle upwelling is also inferred from tomographic images of the velocity structure of the sub-arc mantle¹⁹ and is predicted by some induced-flow thermal models of the mantle wedge³. Petrological arguments have been advanced for mantle upwelling based on inferred primitive melt compositions^{10,20} and inferred low H₂O concentrations²¹.

Although the basaltic melt inclusions can be linked genetically to their host magmas, the origin of the strongly undersaturated melt inclusions is less certain. The lack of relatively elevated incompatible-element contents argues against the strongly undersaturated liquids having formed as low-degree melts of the source that produced the high-Mg basalts. The high normative pyroxene contents of the strongly undersaturated glasses suggest that they formed by low-degree melting of clinopyroxenites or olivine-clinopyroxenites. Pyroxenitic rocks may have formed by percolation of slab-derived fluid or melt and subsequently were entrained and partially melted by upwelling peridotite, or they may have been heated and partially melted by basaltic liquids derived from upwelling peridotite. The high Cl in all of the glass inclusions points to the pervasive presence of traces of a subduction-derived fluid at the onset of melting. The low H₂O/Cl in the inclusion glasses is most consistent with that fluid having been a hypersaline brine that could have carried the subduction elemental signature into the upwelling mantle unaccompanied by substantial amounts of H₂O.

The documented high H₂O contents of some mafic arc magmas^{5,10,12} verify that slab derived fluids are generally involved in melting beneath arcs. However, the discovery of primitive low-H₂O basalts erupted at the volcanic front in the Indonesian and Cascade arcs requires that substantial upwelling and pressure-release melting also take place. This upwelling influences the production of magmas with high or low subduction-fluid contents by augmenting the degree of melting, but the presence of upwelling can only be identified with certainty from the H₂O-poor magmas. If upwelling is widespread and persistent, the shallow sub-arc mantle and lower arc crust are much hotter than widely appreciated. The lowermost arc crust, in particular, may have a temperature near 1,200 °C. In a long-lived arc, the lower arc crust would have been purged of easily fused materials and must be made of highly refractory cumulates or restites. Upwelling would also lead to elevated temperatures in subducting slabs, which could reach temperatures high enough for significant partial melting to occur. □

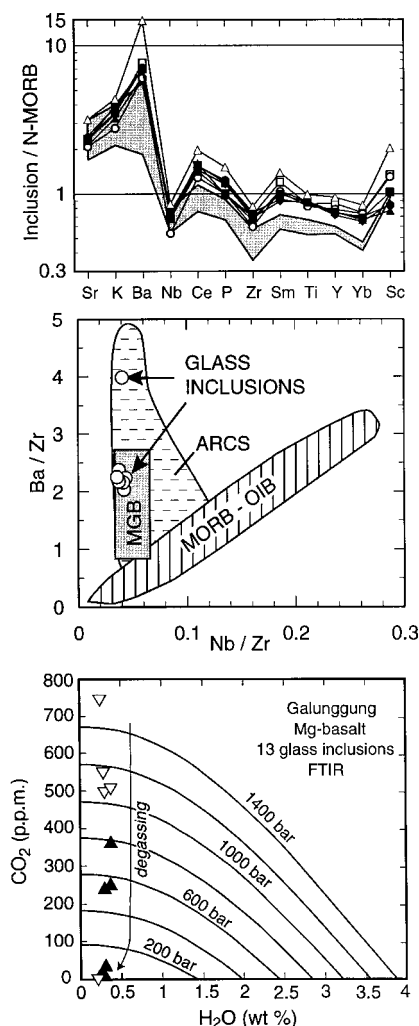


Figure 2 Detailed analyses of Galunggung high-Mg basalts and inclusion glasses. Top panel, MORB-normalized²⁵ trace elements in Galunggung Mg-basalts (shaded) and glass inclusions (symbols: filled, basaltic inclusions; open, undersaturated inclusions). Inclusion trace elements were measured by ion-microprobe, standardized with basaltic glasses. Middle panel, glass inclusions (circles), host Mg-basalts (shaded, MGB), the MORB-ocean island basalt (OIB) array, and other mafic arc magmas²⁶. Bottom panel, glass inclusion H₂O and CO₂ concentrations (symbols: filled, basaltic inclusions; open, undersaturated inclusions) measured by Fourier transform infrared spectroscopy (FTIR), using 3,550 cm⁻¹ and 1,430–1,515 cm⁻¹ absorptions, and extinction coefficients of 63 l mol⁻¹ cm⁻¹ and 375 l mol⁻¹ cm⁻¹ (basaltic glasses) or 355 l mol⁻¹ cm⁻¹ (undersaturated glasses)²⁷; inclusion thicknesses were measured by 0.5-μm-precision dial indicator and Berek compensator.

Received 22 May 1997; accepted 7 January 1998

- Davies, J. H. & Stevenson, D. J. Physical model of source region of subduction zone volcanics. *J. Geophys. Res.* **97**, 2037–2070 (1992).
- Draper, D. S. & Johnston, A. D. Anhydrous PT relations of an Aleutian high-MgO basalt: an investigation of the role of olivine-liquid reaction in the generation of arc high-alumina basalts. *Contrib. Mineral. Petrol.* **112**, 501–519 (1992).
- Furukawa, Y. Depth of the decoupling plate interface and thermal structure under arcs. *J. Geophys. Res.* **98**, 20005–20013 (1993).
- Bacon, C. R., et al. Primitive magmas at five Cascade volcanic fields: melts from hot, heterogeneous sub-arc mantle. *Can. Mineral.* **35**, 397–423 (1997).
- Sisson, T. W. & Layne, G. D. H₂O in basalt and basaltic andesite glass inclusions from four subduction-related volcanoes. *Earth Planet. Sci. Lett.* **117**, 619–635 (1993).
- Sudrajat, A. & Tilling, R. Volcanic hazards in Indonesia: The 1982–83 eruption of Galunggung. *Episodes* **7**, 13–19 (1984).
- Gerbe, M.-C., et al. Mineralogical and geochemical evolution of the 1982–1983 Galunggung eruption (Indonesia). *Bull. Volcanol.* **54**, 284–298 (1992).
- Harmon, R. S. & Gerbe, M.-C. The 1982–83 eruption at Galunggung Volcano, Java (Indonesia): Oxygen isotope geochemistry of a chemically zoned magma chamber. *J. Petrol.* **33**, 585–609 (1992).
- Sobolev, A. V. & Danushevsky, L. V. Petrology and geochemistry of boninites from the north termination of the Tonga Trench: Constraints on the generation conditions of primary high-Ca boninite magmas. *J. Petrol.* **35**, 1183–1211 (1994).
- Sobolev, A. V. & Chaussidon, M. H₂O concentrations in primary melts from supra-subduction zones and mid-ocean ridges: Implications for H₂O storage and recycling in the mantle. *Earth Planet. Sci. Lett.* **137**, 45–55 (1996).
- Dixon, J. E. & Stolper, E. M. An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part II: Applications to degassing. *J. Petrol.* **36**, 1633–1646 (1995).

12. Roggensack, K., Hervig, R. L., McKnight, S. B. & Williams, S. N. Explosive basaltic volcanism from Cerro Negro volcano: influence of volatiles on eruptive style. *Science* **277**, 1639–1642 (1997).
13. Holloway, J. R. & Blank, J. G. Applications of experimental results to C–O–H species in natural melts. *Rev. Mineral.* **30**, 187–230 (1994).
14. Dixon, J. E., Clague, D. A. & Stolper, E. M. Degassing history of water, sulfur, and carbon in submarine lavas from Kilauea volcano, Hawaii. *J. Geol.* **99**, 371–394 (1991).
15. Anderson, A. T. Chlorine, sulfur, and water in magmas and oceans. *Geol. Soc. Am. Bull.* **85**, 1485–1492 (1974).
16. Anderson, A. T. Parental basalts in subduction zones: Implications for continental evolution. *J. Geophys. Res.* **87**, 7047–7060 (1982).
17. Michael, P. Regionally distinctive sources of depleted MORB: Evidence from trace elements and H₂O. *Earth Planet. Sci. Lett.* **131**, 301–320 (1995).
18. Bartels, K. S., Kinzler, R. J. & Grove, T. L. High pressure phase relations of primitive high-alumina basalts from Medicine Lake volcano, northern California. *Contrib. Mineral. Petrol.* **108**, 253–270 (1991).
19. Hasegawa, A., Zhao, D., Hori, S., Yamamoto, A. & Horiuchi, S. Deep structure in the northeast Japan arc and its relationship to seismic and volcanic activity. *Nature* **352**, 683–689 (1991).
20. Tatsumi, Y., Sakuyama, M., Fukuyama, H. & Kushiro, I. Generation of arc basaltic magmas and thermal structure of the mantle wedge in subduction zones. *J. Geophys. Res.* **88**, 5815–5825 (1983).
21. Johnston, A. D. & Draper, D. S. Near-liquidus phase relations of an anhydrous high-magnesia basalt from the Aleutian Islands: Implications for arc magma genesis and ascent. *J. Volcanol. Geotherm. Res.* **52**, 27–41 (1992).
22. Nye, C. J. & Reid, M. R. Geochemistry of primary and least fractionated lavas from Okmok volcano, central Aleutians: Implications for arc magma genesis. *J. Geophys. Res.* **91**, 10271–10287 (1986).
23. Plank, T. & White, W. W. Nb and Ta in arc and mid-ocean ridge basalts. *Eos* **76**, 655 (1995).
24. Tormey, D. R., Grove, T. L. & Bryan, W. B. Experimental petrology of normal MORB near the Kane Fracture Zone: 22°–25° N, mid-Atlantic ridge. *Contrib. Mineral. Petrol.* **96**, 121–139 (1987).
25. Pearce, J. A. in *Andesites* (ed. Thorpe, R. S.) 525–548 (Wiley, New York, 1982).
26. Leeman, W. P., Smith, D. R., Hildreth, W., Palacz, Z. & Rodgers, N. Compositional diversity of late Cenozoic basalts in a transect across the southern Washington Cascades: Implications for subduction zone magmatism. *J. Geophys. Res.* **95**, 19561–19582 (1990).
27. Ihinger, P. D., Hervig, R. L. & McMillan, P. F. Analytical methods for volatiles in glasses. *Rev. Mineral.* **30**, 67–122 (1994).

Acknowledgements. We thank R. Tilling and S. Weaver for samples from Galunggung; J. Lowenstein for instruction in FTIR spectroscopy; G. Layne for performing ion-microprobe analyses; and C. Bacon, M. Clyne, T. Crawford and T. Grove for reviews that helped us to clarify our arguments. This work was supported by the USGS Volcano Hazards Program.

Correspondence and requests for materials should be addressed to T.W.S. (e-mail: tsisson@usgs.gov).

Foot posture in a primitive pterosaur

J. M. Clark*, J. A. Hopson†, R. Hernández R.‡, D. E. Fastovsky§ & M. Montellano‡

* Department of Biological Sciences, George Washington University, Washington DC 20052, USA

† Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA

‡ Instituto de Geología, Universidad Nacional Autónoma de México, Del. Coyoacan, DF 04510, Mexico

§ Department of Geology, University of Rhode Island, Kingston, Rhode Island 02881, USA

The nature of the hindlimb posture and gait of pterosaurs has been controversial^{1–16}, partly because most of the pterosaur skeletons that have been found were flattened in thin-bedded rocks, therefore obscuring three-dimensional anatomy. A major controversy concerns the extent to which pterosaurs move on the ground; they have been variously interpreted as ranging from sprawling, quadrupedal walkers to erect, bird-like bipedal cursors¹. Study of pelvis and femur material from the derived group Pterodactyloidea^{11–13} has resolved which movements are possible at the hip, but the lack of three-dimensional, articulated pterosaur feet has prevented examination of all of the movements that are possible within the foot. We have found a large, uncrushed, partial skeleton of a new species of the basal pterosaur *Dimorphodon* in thick-bedded deposits of Tamaulipas, Mexico; this material includes such a three-dimensional foot. The nature of this skeleton contradicts an important part of the cursorial interpretation, that is, that only the toes contacted the ground during terrestrial locomotion^{2–7}. The flattened metatarsal-phalangeal joint at the base of the first four toes of this specimen would not allow such a digitigrade posture without separating

most of the joints. A flat-footed stance is consistent with presumed footprints of pterosaurs^{8–10} that show impressions of the entire sole of the foot.

Monophyletic hierarchy⁶: Diapsida, Archosauria, Ornithodira, Pterosauria,

Dimorphodontidae

Dimorphodon weintraubi, sp. nov.

Etymology. *weintraubi*: after the late Dr Robert L. Weintraub.

Holotype. Universidad Nacional Autónoma de México Instituto Geología de México (IGM) 3494, an articulated partial skeleton including the posterior part of the skull and first four cervical vertebrae, left and right scapulocoracoids, left humerus, second phalanx of left wing finger, right wing distal to mid-humerus, and right leg distal to mid-tibiotarsus.

Locality and horizon. Lower part of the La Boca Formation in Huizachal Canyon, Tamaulipas, Mexico, from a 1-metre-thick volcanoclastic siltstone deposited as a subaerial, waterlain debris flow^{17,18}. The fauna indicates a late Early or early Middle Jurassic age¹⁷.

Diagnosis. *D. weintraubi* differs from *D. macronyx* (Buckland 1829), the only other species in the genus¹⁹, in that its first wing phalanx is not significantly shorter than the ulna¹⁴. A sesamoid at the base of pedal claws, reported for *D. macronyx*¹⁴, is also not seen. The holotype is larger than the largest specimens of

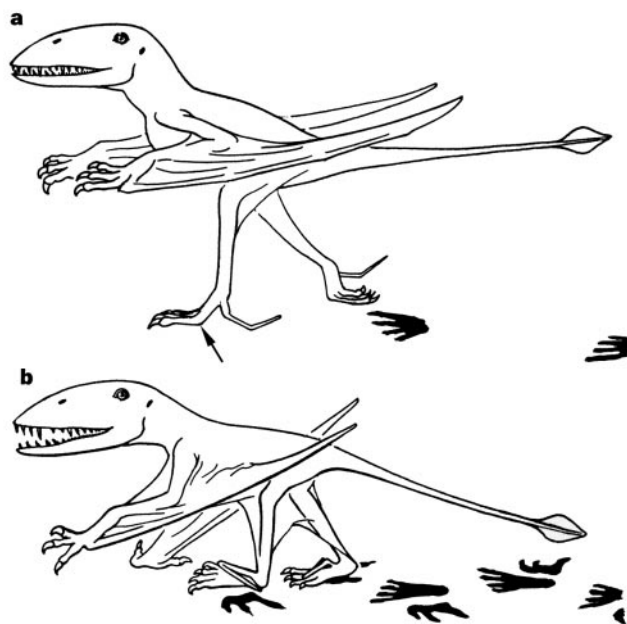


Figure 1 Alternative reconstructions of locomotor posture and trackways of *D. macronyx*. **a**, *D. macronyx* reconstructed as a bipedal digitigrade runner (redrawn from ref. 13, adapted from ref. 2) in which substantial hyperextension occurs at the metatarsal-phalangeal (MP) joints (arrow) when the foot contacts the ground. **b**, *D. macronyx* as a quadrupedal, plantigrade walker, in which the entire plantar surface of the foot contacts the ground and only slight hyperextension occurs at the MP joints at the end of the propulsive phase. In **a**, the hindlimb moves in a parasagittal plane and the footprints are close to the animal's midline, as dictated by the bipedal-running hypothesis. In **b**, the forelimbs and hindlimbs may have been more sprawling than indicated here and the footprints may have been further from the midline. The reconstruction of the flight membrane in **b** is based on the interpretation of *Sordes pilosus*¹⁶; the depiction of the folding of the fifth pedal digit over the foot is based on preservation seen in *D. weintraubi* and other primitive pterosaurs^{14,23,24}. The footprints in **a** are hypothetical: the spacing and orientation of the footprints in **b** are based on a trackway of *Pteranichnus cf. saltwashensis*⁹.

D. macronyx, and is the largest known nonpterodactyloid pterosaur; its 60.5-mm-long metacarpal IV is longer than the 57.5-mm element of the largest previously reported basal pterosaur²⁰.

This species is similar to *D. macronyx* from the Lower Jurassic of England^{2,14,19} in nearly all skeletal features known for both taxa. Although the cranial material of *D. macronyx* and *D. weintraubi* has few elements in common, the unusually deep, vertical basipterygoid processes of the basiphenoid of *D. weintraubi* indicate that the pterygoid and quadrate extended well below the occipital condyle, and that the temporal region was similar to that of *D. macronyx* in being significantly deeper than in other pterosaurs. In both taxa, the second phalanx of pedal-digit V is unusually long and straight²¹, and the first pedal digit diverged medially at least 15° (although the distribution of these features among pterosaurs is incompletely known). *Dimorphodon* is at present considered to be a basal pterosaur that is more distantly related to pterodactyloids than are some other 'rhamphorhynchoid'-grade taxa^{21,22}.

Specimens of *D. macronyx* have been important in interpreting primitive pterosaurs as highly cursorial bipeds², and were the only specimens directly cited in support of digitigrady³. Hindlimbs were proposed to have moved with an erect, parasagittal gait, and with a digitigrade foot posture (Fig. 1). In a fully digitigrade gait, the ankle and metatarsus are raised above the ground and only the digits contact the substrate. It was argued^{2,3}, on the basis of comparisons with birds and dinosaurs, that dorsiflexion (hyperextension) occurred at the metatarsal-phalangeal (MP) joints of digits I–IV. The shape of these joints² and the symmetry of the foot on either side of the long third toe³ were used to support this hypothesis.

The MP joints of digits I–IV of *D. weintraubi* (Figs 2 and 3) indicate, however, that little rotation, far less than in birds, was possible. The articular surfaces on the distal end of the avian tarsometatarsus (Figs 3d and 4) extend in a broad arc over 180° in lateral view. This surface forms a trochlea (a pulley-shaped structure) that articulates with a ridge on the proximal phalanx. In *D.*

weintraubi, however, the articular surface on the distal end of each metatarsal extends to only a limited extent onto the dorsal surface of the shaft (Figs 3a, e and 4), and a complementary groove and ridge are absent from the articular surfaces. In *D. macronyx*, a broad groove present on the ventral surface of the distal end of metatarsals I–IV was interpreted to be a trochlea², but in *D. weintraubi* and *D. macronyx* (Natural History Museum, London) the proximal phalanx does not articulate within this groove, and it was presumably occupied by flexor tendons that were inserted on more distal phalanges.

Most important, the distal joint surfaces on the metatarsals of *D. weintraubi* are only slightly convex and those of the opposing phalanges only slightly concave (Figs 2 and 3b, c and e), forming an unusually flat MP joint compared with those in members of the two groups of living archosaurs. When articulated with the joint surfaces flush, the shaft of each phalanx of *D. weintraubi* would be directly in line with that of the metatarsus (except for the first toe, which diverged medially). Even a small amount of hyperextension of the phalanx on the metatarsal (as little as 5°) would disarticulate a large portion of the joint surface (Fig. 3e).

When the foot is considered as a whole, the interphalangeal joint surfaces also allow little hyperextension. The joint between the ungual and penultimate phalanx in the three digits in which the ungual is preserved allows for no hyperextension, indicating that an unguligrade posture (walking on unguals) was impossible. The remaining interphalangeal joints allow differing degrees of dorsiflexion among the digits, with the amount of dorsiflexion increasing from medial to lateral digits, as in plantigrade reptiles but not in birds. In digit I, the single interphalangeal joint allows for no hyperextension, whereas the four interphalangeal joints of digit IV together allow for elevation of the metatarsus approximately 50° above the ground (assuming its claw to be similar to those of other digits). As digit I is deflected medially by its oblique MP joint, it may not have reached the ground in a digitigrade posture. However, a

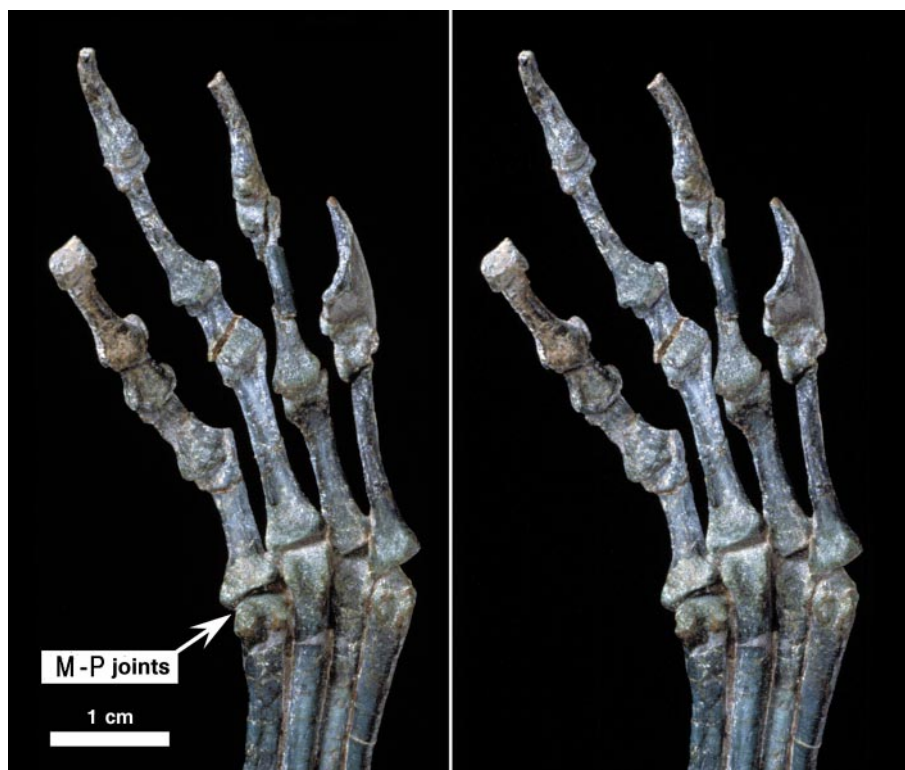


Figure 2 Stereophotographs of the right foot of *D. weintraubi* in ventral view. The toes have been hyperextended and dislocated slightly to the right (left side of the photograph), exposing the flat articulation surfaces on both sides of the

metatarsal-phalangeal joints (MP Jts.). Digit IV, without a claw, is slightly flexed; when extended it would have reached beyond the level of the end of the preungual phalanges of digit III.

limit to digital hyperextension on the medial side of the foot is set by digit II, which was capable of a maximum of 20° of dorsiflexion. A parasagittal stride would begin with the digits parallel to the ground, and maximum dorsiflexion would have occurred at the end of the stride. This limited degree of hyperextension on the inner side of the foot is not consistent with a wholly digitigrade, parasagittal stride, in which digits II–IV support the weight of the body throughout the contact phase of the foot.

It has been argued³ that the symmetry of the pterosaur pes, with the third digit being the longest, supports the idea that pterosaurs moved rapidly and displayed obligate bipedality. However, the pes is not symmetrical in *D. weintraubi*. Assuming that the missing claw of digit IV is as long as that of digit III, digit IV is longer than digit III. Although the third toe is the longest in pterodactyls and rhamphorhynchids²³, in the basal pterosaurs *Peteinosaurus*²⁴ and *Eudimorphodon*¹⁵ and in *D. macronyx*¹⁴ (J.A.H., unpublished observations) the toes increase in length from digit I to digit IV. Thus, the digits of basal pterosaurs have proportions that are similar to those of the sprawling feet of ambulatory, plantigrade reptiles rather than those of the cursorial, digitigrade feet of bipedal reptiles.

The phalanges of the slender, elongated fifth digit, to which the wing membrane may have been attached¹⁶, apparently did not participate in supporting the foot. The robust metatarsal is positioned posterolaterally to the other metatarsals, and its plantar surface would contact the substrate in a plantigrade stance. How-

ever, the nature of the MP joint does not indicate that the proximal phalanx was capable of projecting much further ventrally than metatarsals I–IV, and its length and slenderness would have provided little mechanical advantage in supporting the rest of the foot.

Other features of digits I–IV of the *D. weintraubi* foot indicate a capacity for grasping that is consistent with an ability to climb but is unexpected in an obligate cursor. The claws are moderately curved (nearly as strongly as the claws of the manus); all phalanges except the most proximal have well developed flexor tubercles for the insertion of digital flexors (Fig. 2); and all of the IP joints allow for extensive flexion of the digits (as exhibited by digit IV; Fig. 2). Furthermore, the phalangeal proportions of the digits of *Dimorphodon* and other basal pterosaurs are similar to those of birds with grasping feet (that is, perching, climbing, and raptorial species) and unlike those of primarily ground-living birds, bipedal dinosaurs and the primitive dinosauroforms *Lagerpeton* and *Marasuchus* (Fig. 5).

The similarity of the foot of *D. weintraubi* to that of other basal pterosaurs^{15,24} suggests that the features outlined here are primitive features of the group. The lack of cursorial adaptations in the foot of *D. weintraubi* contradicts the reconstruction of basal pterosaurs as rapidly moving, digitigrade cursors^{2,3}. On the contrary, the capacity for grasping exhibited by this foot and the similarity of its phalanges to those of birds with grasping feet indicates that basal pterosaurs may have been scansorial and, perhaps, arboreal^{1,14,15}. □

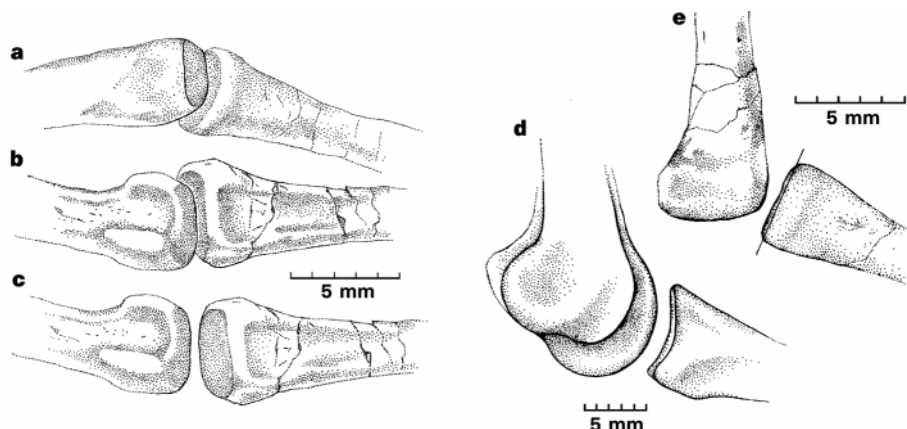


Figure 3 Metatarsal-phalangeal (MP) joints of *D. weintraubi* and a bird. **a–c**, MP joints of *D. weintraubi*. **a**, Digit II in dorsal view; **b**, digit III in ventral view; **c**, reconstruction of digit III in ventral view when separated. **d**, Distal end of tarsometatarsus and proximal end of proximal phalanx of digit III when hyperextended in a bird (albatross), showing the rounded trochlea extending onto the tarsometatarsal shaft and the concave proximal end of the phalanx.

e, Reconstruction of a similarly hyperextended articulation of metatarsal and proximal phalanx of digit III of *D. weintraubi*, separated slightly, showing consequent disarticulation of most joint surfaces dorsal and ventral to the small area of contact. The line through the phalanx shows the deepest extent of the articulating surface.

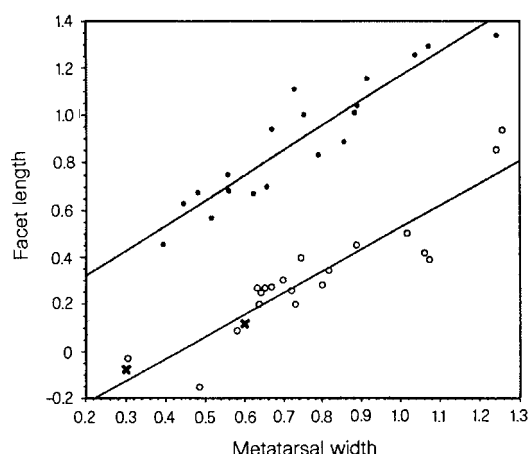


Figure 4 Length of the articulation surface on the distal end of the metatarsus in avian and non-avian reptiles, illustrating the greater extent of this surface in birds. The graph plots the log of the length of the distal articulation facet on the dorsal surface of metatarsal III (y-axis), measured parallel to the shaft from the distal end of the bone to the proximodorsal end of the articulation surface, against the log of the distal width of metatarsal III as a measure of size (x-axis). The short dorsal articular surface in both *D. weintraubi* and *D. macronyx* is comparable to that in reptiles that do not hyperextend their toes and are plantigrade rather than digitigrade as birds are. Filled circles, birds; open circles, lizards, crocodylians, and turtles; crosses, *D. macronyx* (left) and *D. weintraubi* (right). For reptiles, $(\log y) = 0.93(\log x) - 0.406$, $r^2 = 0.827$; for birds, $(\log y) = 1.056(\log x) + 0.111$, $r^2 = 0.865$.

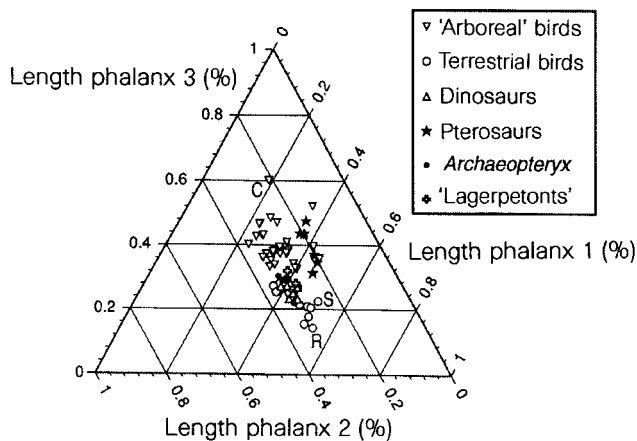


Figure 5 Ternary diagram showing relative lengths of the three non-ungual phalanges of pedal digit III in living birds and fossil archosaurs expressed as a percentage of the summed lengths of the three phalanges. For example, the phalangeal proportions of the grasping digit III of a swift (*Chaetura cinereiventris*), represented by the inverted triangle (identified as C) closest to the apex of the diagram, are, from first to third phalanges: 19, 21, and 60%. Conversely, the phalangeal proportions of the fully terrestrial, and highly cursorial, rhea (*Rhea americana*), represented by the open circle closest to the base of the diagram (R), are 54, 32, and 14%. We include data from studies of 32 climbing, perching, or raptorial birds (3 woodpeckers, a toucan, a colie, 4 parrots, 2 cuckoos, a hoatzin, 4 owls, 6 falconiforms, an oilbird, a hornbill, a swift, and 7 perching passeriforms), all of which possess grasping feet. Data from studies of 22 terrestrial birds (4 ratites, a tinamou, 3 penguins, 4 galliforms, 2 herons, a stork, an ibis, 3 charadriiforms, a sand grouse, a cuckoo (*Geococcyx californicus*, the roadrunner), and a falconiform (*Sagittarius serpentarius*, the secretary bird labelled S) are included. Ten theropods and one ornithomimid are the dinosaurs included. Two specimens of the early fossil bird *Archaeopteryx* are represented by filled circles and two primitive dinosauriforms are represented by open crosses (*Marasuchus*, upper left, and *Lagerpeton*). The lower cluster of three pterosaurs (stars) includes (lowest to highest stars) *D. macronyx*, *D. weintraubi*, and *Peteinosaurus zambellii*; the upper cluster includes (left to right) *Rhamphorhynchus muensteri*, *R. longicauda*, and *Anurognathus ammoni*. The pterosaur points lie on or close to a cluster of five non-terrestrial falconiforms (the most dorsal, above *A. ammoni*, being the osprey, *Pandion haliaetus*).

Received 6 May; accepted 26 November 1997.

- Unwin, D. Pterosaur locomotion: joggers or waddlers? *Nature* **327**, 13–14 (1987).
- Padian, K. Osteology and functional morphology of *Dimorphodon macronyx* (Buckland) (Pterosauria: Rhamphorhynchoidea) based on new material in the Yale Peabody Museum. *Postilla* **189**, 1–44 (1983).
- Padian, K. A functional analysis of flying and walking in pterosaurs. *Paleobiology* **9**, 218–239 (1983).
- Padian, K. in *Biomechanics in Evolution* (eds. Sereno, J. M. V. & Wootton, R. J.) 145–160 (Cambridge Univ. Press, 1991).
- Padian, K. & Rayner, J. V. The wings of pterosaurs. *Am. J. Sci.* **239-A**, 91–166 (1993).
- Sereno, P. Basal archosaurs: phylogenetic relationships and functional implications. *J. Vert. Paleontol.* **11**, (suppl. to no. 4) 1–53 (1991).
- Padian, K. in *Short Papers Third Symp. Mesozoic Terrestrial Ecosystems* (eds. Reif, W.-E. & Westphal, F.) 163–168 (Attempto, Tübingen, 1984).
- Mazin, J., Hantzpergue, P., Lafaurie, G. & Vignaud, P. Des pistes de ptérosaures dans le Tithonien de Crayssac (Quercy, France). *C.R. Acad. Sci. II, Sci. Terr. et Planet* **321**, 417–424 (1995).
- Lockley, M. G. et al. The fossil trackway *Pterichnus* is pterosaurian, not crocodilian: implications for the global distribution of pterosaur trackways. *Ichnos* **4**, 7–20 (1995).
- Bennett, S. C. Terrestrial locomotion of pterosaurs: a reconstruction based on *Pterichnus* trackways. *J. Vert. Paleontol.* **17**, 104–113 (1997).
- Wellnhofer, P. Terrestrial locomotion in pterosaurs. *Hist. Biol.* **1**, 3–16 (1988).
- Bennett, S. C. A pterodactylid pterosaur from the Santana Formation of Brazil: implications for terrestrial locomotion. *J. Vert. Paleontol.* **10**, 80–85 (1990).
- Wellnhofer, P. *The Illustrated Encyclopedia of Pterosaurs* (Crescent, New York, 1991).
- Unwin, D. New remains of the pterosaur *Dimorphodon* (Pterosauria: Rhamphorhynchoidea) and the terrestrial ability of early pterosaurs. *Mod. Geol.* **13**, 57–68 (1988).
- Bennett, S. C. The arboreal leaping theory of the origin of pterosaur flight. *Hist. Biol.* **29**, 373–386 (1997).
- Unwin, D. M. & Bakhrina, N. N. *Sordes pilosus* and the nature of the pterosaur flight apparatus. *Nature* **371**, 62–64 (1994).
- Clark, J., Montellano, M., Hopson, J. & Fastovsky, D. in *In the Shadow of the Dinosaurs* (eds. Fraser, N. & Sues, H.-D.) 295–302 (Cambridge, New York, 1994).
- Fastovsky, D. et al. Depositional environments of a Middle Jurassic terrestrial vertebrate assemblage, Huizachal Canyon, Mexico. *J. Vert. Paleontol.* **15**, 561–575 (1995).
- Owen, R. *Monograph of the Fossil Reptilia of the Liassic Formations I. Part III* (Palaeontographical

Society, London, 1870).

- Galton, P. A rhamphorhynchoid pterosaur from the Upper Jurassic of North America. *J. Paleontol.* **55**, 1117–1122 (1981).
- Kellner, A. W. *Description of New Material of Tapejaridae and Anhangueridae* (Pterosauria, Pterodactyloidea) and Discussion of Pterosaur Phylogeny. Thesis, Columbia Univ. (1996).
- Unwin, D. in *Short Papers Sixth Symp. Mesozoic Terrestrial Ecosystems and Biota* (eds. Sun, A. & Wang, Y.) 69–72 (China Ocean, Beijing, 1995).
- Wellnhofer, P. *Pterosauria. Handbuch der Paläohierpetologie* **19** (Fischer, Stuttgart, 1978).
- Wild, R. Die flugsaurier (Reptilia, Pterosauria) aus der oberen Trias von Cene bei Bergamo, Italien. *Boll. Soc. Paleol. Ital.* **17**, 176–256 (1978).

Acknowledgements. This work was supported by the National Geographical Society and National Science Foundation. We thank the Huizachal Ejido and the Consejo Estatal para la Cultura y las Artes de Estado de Tamaulipas for permission to collect fossils in Huizachal Canyon; the Instituto de Geología de U.N.A.M. for support; the Instituto Nacional de Antropología e Historia for permission to borrow the specimen; C. Vanderslice, D. Chaney, and A. Davidson for fossil preparation; M. Ellison, M. Parrish and C. Vanderslice for illustrations; L. Grande and those who helped collect the fossil; M. Lockley, K. Padian, S. Gatesy, R. Blob, M. Carrano, F. A. Jenkins Jr., J. Gauthier, and S. Swartz for comments and the Field Museum of Natural History, the National Museum of Natural History, and the Natural History Museum, London for access to specimens.

Correspondence and requests for materials should be addressed to J.M.C. (e-mail: jclark@gwis2.circ.gwu.edu).

Explaining the geographic distributions of sexual and asexual populations

Joel R. Peck, Jonathan M. Yearsley & David Waxman

Centre for the Study of Evolution, University of Sussex, Brighton BN1 9QG, UK

Examination of the geographic distributions of sexual organisms and their asexual, or parthenogenetic, competitors reveals certain consistent patterns. These patterns are called geographic parthenogenesis^{1–8}. For example, if we compare sexual organisms with closely related asexuals, we find that, in the Northern Hemisphere, there is a strong tendency for the asexuals to occur further to the north. One researcher to document this pattern is Bierzychudek, who examined 43 cases (drawn from 10 genera) where the geographic distributions of a sexual plant and a closely related asexual are known⁴. In 76% of these cases, the asexual plant's range was more northerly than the range of the sexual. Some of the remaining cases probably fit with this pattern, but more data must be obtained before this suggestion can be confirmed. Asexuals also tend to occur at high altitudes, and in marginal, resource-poor environments^{1–8}. We have constructed a mathematical model of a habitat that stretches from south to north in the Northern Hemisphere. Our computer simulations based on this model support the idea that a single basic process may account for much of what is known about geographic parthenogenesis. This process involves the movement of individuals from areas in which they are well adapted to areas where they are poorly adapted.

Details of our model are given in Box 1. We assume that the hypothetical organisms living in the habitat are obligately sexual hermaphrodites. Each adult produces a large number of offspring through 'female effort' (for example, by producing seeds or eggs), and mating is random within local areas. We also assume that the growing season is longer as one moves further south, and so, over the course of a generation, adults in southern regions produce more offspring through female effort than do adults in northern regions. Within a local area, the expected number of offspring produced by female effort is the same for all adults.

After the production of offspring the adults die, and the juveniles are subject to viability selection. A given juvenile's chance of surviving this stage depends on its phenotype, and on the local environment within which it is born. The optimal phenotype varies from one location to another. We assume that phenotype is entirely determined by genotype. Offspring phenotypes are normally distributed around the mean of the phenotypes of their parents (this is the 'infinitesimal model'⁹).

After viability selection, some juveniles migrate to other regions.

The direction and distance that a juvenile migrates are chosen at random. We assume that the local population density of adults that can be sustained by the available resources is the same everywhere within the habitat. Thus, after migration, randomly selected juveniles die at such a rate as to equalize the population density in all areas. The remaining juveniles mature to adulthood, and give birth to the next generation.

Most multicellular asexual organisms are thought to have been derived, in the relatively recent past, from sexual ancestors^{7,10}. We therefore considered a situation where the habitat is occupied initially by an obligately sexual population at equilibrium. We then selected an adult at random from this population, and assumed that this adult has a newly arisen mutation that makes her entirely asexual, and that causes her to produce only asexual offspring. Thus the initial mutant and all of her descendants reproduce exclusively through female effort. The expected fertility of an asexual adult is assumed to be proportional to the expected number of offspring produced by female effort by sexual adults in the same region. Let K

represent this latter quantity in a particular region. We assume that asexual adults in the same region produce an average of ΘK offspring each, where $\Theta > 0$; the value of Θ is the same for all asexual adults.

In general, we can expect to find that sexuals and asexuals living in the same region will not have equal fertility (that is, we can expect $\Theta \neq 1$). Asexuals do not have to find mates, and so they may be more efficient than sexuals, leading to $\Theta > 1$. However, asexuality can produce problems, such as a high rate of mortality from deleterious mutations and a high level of susceptibility to parasites^{7,10–20}. Factors such as these may lead to depressed fertility in asexuals (so $\Theta < 1$). For each simulation trial, we chose the value of Θ at random. (The other parameters of the model were also chosen at random, as described in Box 1).

We assume that all descendants of the initial asexual mutant have the same phenotype as their ancestor until a mutation occurs among them that changes the phenotype. We refer to this collection of phenotypically identical descendants as a clone. The number of

Box 1

For the purposes of the simulations, it is convenient to 'discretize' both the environment and the possible phenotypes. The environment was divided into an infinite number of areas, numbered sequentially from $-\infty$ in the far south to ∞ in the far north. The population is inviable in all areas except for those numbered $1, 2, \dots, \Omega$ (see below for details). For the main set of simulations, $\Omega = 11$. However, we experimented with larger values of Ω , and this had no qualitative effect on the results.

Let the expected number of offspring produced by a sexual adult through female effort in the southern-most habitable area (area number 1) be denoted by K_1 . The same statistic in area i is denoted by K_i . We assume that K_i is given by $K_i = K_1[1 - (2i - 2)/(\Omega - 3)]$ for $1 \leq i \leq \Omega$. Thus fertility undergoes a three-fold decrease from the most southern to the most northern habitable area.

We assume that phenotypes are discrete, and the possible phenotypic values (denoted by z) range sequentially by integers from $-\infty$ to ∞ . If the mean phenotypic value of a pair of parents is given by $\bar{z}$, the probability that the phenotype of a particular offspring takes on an integer value i namely $V(i|\bar{z})$ is given by the following 'discretized' normal distribution: $V(i|\bar{z}) = Q \exp(-(i - \bar{z})^2/\beta_z)$, where β_z is a positive constant. Here, Q is chosen to ensure that $\sum_{i=-\infty}^{\infty} V(i|\bar{z}) = 1$.

After birth, viability selection occurs. In any given area i , the probability that an individual with phenotype z will survive viability selection is denoted by $W(i, z)$. For $1 \leq i \leq \Omega$, $W(i, z)$ is given by the traditional 'non-optimal' function: $W(i, z) = \exp(-(z - i)^2/\beta_v)$, where $\beta_v > 0$. Thus the optimum phenotype in a given area i is given by $z = i$. We assume that $W(i, z) = 0$ for $i \leq 0$ or $i > \Omega$.

During the migration phase, the probability that an individual living in area i migrates to area j , namely $M(i, j)$ is given by $M(i, j) = \tanh(1/(2\beta_m)) \exp(-|i - j|/\beta_m)$, where $\beta_m > 0$. Thus the probability that an individual remains in the area where it was born is given by $\tanh(1/(2\beta_m))$. The average absolute distance (in terms of values of i) that an individual moves is $1/\sinh(1/\beta_m)$. The probability that an individual in area i will migrate beyond $x = 1$ or $x = \Omega$ is equal to $\exp(-\Omega/(2\beta_m)) \cosh([2i - \Omega - 1]/(2\beta_m)) / \cosh(1/(2\beta_m))$.

We began simulations with a sexual population in which phenotypes $z = 1, z = 2, \dots, z = \Omega$ were present in areas $1, 2, \dots, \Omega$ (the initial densities were randomized in each area). We then ran the simulation until equilibration, defined as the point when the average postmigration density of the population (averaging over areas $1, 2, \dots, \Omega$) was changing by less than 0.01% over the course of 100 generations.

To introduce the first asexual mutant into the equilibrium population, our programme used a procedure that, mathematically speaking, is exactly equivalent to the following sequence of events. First, a sexually produced adult is chosen at random, and its location and phenotype are noted. We assume that this adult is asexual. Then previously published methods²⁵ are used to calculate the probability (P) that a clone derived from this adult will become established (that is, it will not be lost from the population as a result of stochastic processes). If the clone is established (which happens with

probability P) then it grows in number until a new equilibrium is reached. If the clone is not established, a new sexually produced adult is selected to begin the procedure again. This continues until a clone becomes established.

It can be shown that, after a clone becomes established, the proportion of members of the clone that reside in each area will eventually approach a particular distribution (an eigenvector), and this invasion distribution will remain for as long as members of the clone are very rare in all areas. The invasion distribution depends on the phenotype of members of the clone, but is independent of the area in which the initial asexual mutant appeared. The eigenvector in question is calculated from an $\Omega \times \Omega$ matrix, in which the entry in the i th row and the j th column is given by $\Theta M(i, j) K(i) W(i, z^*) \bar{N}(i)$, where z^* gives the phenotype of the new clone and $\bar{N}(i)$ is the equilibrium density (before the introduction of the new clone) of all individuals (regardless of whether sexual or asexual) on site i after viability selection and migration.

To calculate the postinvasion equilibrium distribution for asexuals and sexuals, we added asexual juveniles having the phenotype of the newly established clone to the juveniles living in each of the areas, immediately after viability selection and migration. The relative numbers of members of the new clone in each area is determined by the invasion distribution above. We normalized the densities of the new clone to ensure that, initially, the ratio of clone density to total density could not exceed 10^{-3} in any area. In general, we allowed the population to evolve until the average postmigration density of both species (averaging over areas) was changing by less than 0.01% over the course of 100 generations. Results from numerical experiments strongly suggested that this criterion was usually sufficiently stringent to ensure accurate calculation of the equilibrium. However, for some trials in which evolution appeared to proceed very slowly, we applied an even more stringent criterion for determining when to stop the simulation.

After the first clone to invade successfully has reached equilibrium, we initiate a series of new mutants as follows. For each new mutant, the probability that the mother of the mutant will have a particular phenotype is equal to the frequency of that phenotype among all the asexuals. The phenotype of the new mutant is chosen using a discretized normal distribution. The probability that the mutant will have phenotype i is given by $C \exp(-(i - j)^2/2)$, where j is the phenotype of the parent of the mutant, and C is chosen to ensure that the distribution sums to unity when the sum is over all i such that $-\infty \leq i \leq \infty$. After each new mutant is produced, it either becomes established, leading to a new equilibrium, or it goes extinct. This continues with successive mutants until an equilibrium is achieved for which there is no possible phenotype that is not present among the asexuals living in the habitat, and that could possibly become established (that is, for which $P > 0$).

Before each trial, we choose Θ, β_v, β_m and β_m from rectangular distributions. The intervals for these distributions were as follows: for Θ , [0.7, 1.1]; for β_v , [0.01, 0.5]; for β_m , [1, 10]; and for β_m , [1, 3].

living individuals who are part of a clone will either become large (in which case we say that the clone is established), or it will fall to zero (which means extinction of the clone). If the clone becomes extinct, we choose another sexual adult and begin again. If the number becomes large, we allow a new equilibrium to develop. Once an asexual clone is established and a new equilibrium has developed, we initiate a new clone by assuming that an adult asexual undergoes a mutation that alters its phenotype, and then we allow evolution to proceed until this new clone becomes extinct or becomes established and a new equilibrium develops. We continue with this process of introducing new asexual mutants until there is no longer any chance that a new asexual clone can become established. Analytic methods are used to determine when this final distribution has been achieved.

We ran 2,500 computer trials using these procedures. In 25.5% of these trials, we found that only asexuals were present by the time the trial ended. This happened whenever $\Theta > 1$ (so that asexuals had a fertility advantage), and sometimes when $\Theta < 1$ (in 0.5% of the cases where sexuals were absent at the end of the trial, $\Theta < 1$). In 0.1% of the 2,500 trials, only sexuals were present at the final equilibrium. Finally, in 74.4% of the trials, both sexuals and asexuals were present at the final equilibrium; we call these the polymorphic trials.

In 96% of the polymorphic trials, we found that, once the final

equilibrium had been reached, the average position of sexual adults was further to the south than the average position of asexual adults. This outcome corresponds to the usual finding in nature^{1,2,4-8}. We also found that, in 94% of the polymorphic trials, the density of sexual adults decreased from south to north, and the density of the asexual adults increased from south to north. An example of this sort of outcome is shown in Fig. 1a. In 4% of the polymorphic trials the final average position of the sexuals was further to the north than that of the asexuals (Fig. 1b, c).

Why are patterns similar to those found in nature so common among the polymorphic trials? To answer this question, let us consider, in more detail, the trial for which the final distribution is given in Fig. 1a. In this trial (as in all the trials) the population density of sexual adults was the same everywhere before the first successful invasion of an asexual clone. However, data from the simulations show that, at this initial equilibrium, juveniles were least viable in the north. This is a consequence of the relatively high levels of fertility in the southern populations. As a result of the fertility differences, migrants and their recent descendants are more common (in terms of proportions) in northern populations than in southern ones. As migrants tend to have phenotypes that are not well adapted for local conditions, this depresses viability in northern populations.

Because of blending inheritance, the deleterious effects for sexuals

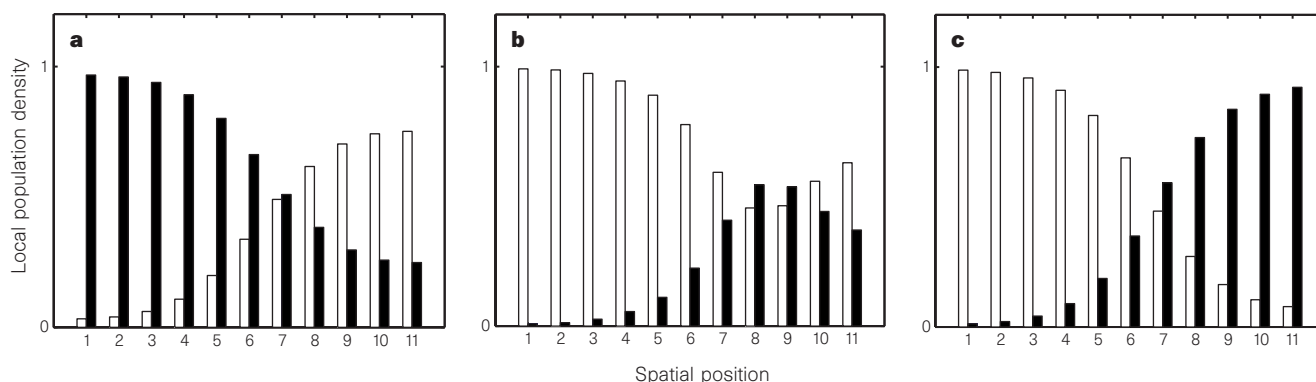


Figure 1 Final equilibrium density distributions for three simulation runs. The abscissas give spatial position, with the southern-most position being farthest to the left. The ordinates are proportional to local population density. Black bars give data for sexual individuals, and white bars for asexuals. In 96% of the polymorphic trials, the average position for the sexuals was further to the south than for the

asexuals, once the final equilibrium had been reached. An example is shown in **a**. In 4% of the polymorphic trials, the average position of the asexuals was further south at the final equilibrium, as shown in **b** and **c**. In **a**, the values of Θ , β_s , β_v and β_m were 0.743, 0.140, 2.27 and 2.67, respectively. In **b**, these same values were 0.961, 0.028, 2.99 and 1.63. In **c**, the values were 0.935, 0.461, 1.28 and 1.17.

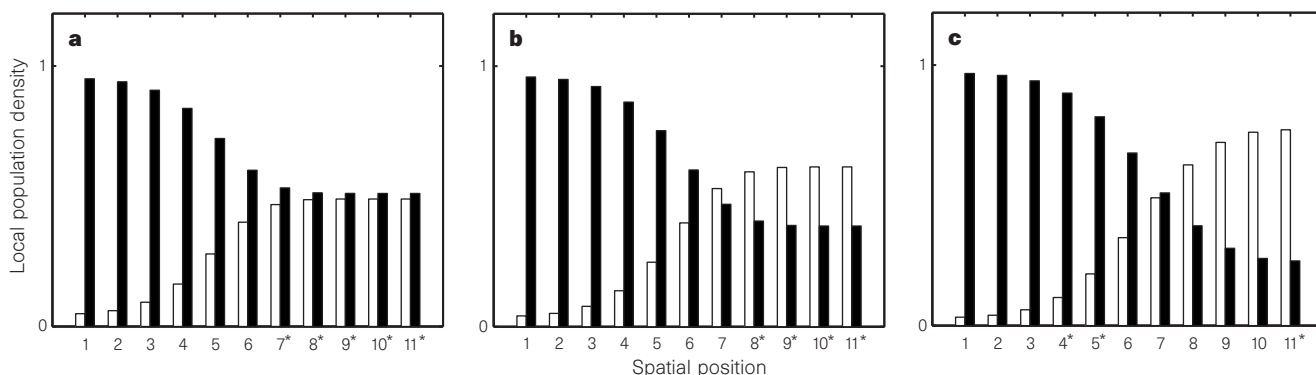


Figure 2 The intermediate equilibria leading to the final equilibrium shown in Fig. 1a. The abscissas give spatial position, and the ordinates are proportional to local population density. Black bars give data for sexual individuals, and white bars for asexuals. The asterisks indicate that an asexual clone with the optimal phenotype for the area marked had a non-zero probability of invading successfully. **a**, The equilibrium established after one successful invasion (with

an asexual clone optimal for area 6). **b**, The equilibrium after the second successful invasion (with an asexual clone optimal for area 7). **c**, The equilibrium after three more successful invasions (with clones optimal for areas 9, 8 and 10, in that order). An additional three successful invasions (by clones optimal for areas 5, 11 and 4) occurred before establishment of the equilibrium shown in Fig. 1a.

of migration into a particular region can be eliminated, in part, if similar numbers of immigrants come from the north and the south. However, if more immigrants come from the south, the mean phenotype in a given region will tend to be pushed to a value that is optimal for a more southerly region. It can be shown that the fertility of a local population, in comparison with the fertility of the population immediately to its south, becomes smaller (in proportional terms) as one moves further north. Thus we have another reason why the loss of fitness resulting from migration tends to be largest in the far north.

In the trial shown in Fig. 1a, $\Theta < 1$, so asexuals have an intrinsic fertility disadvantage. Analysis shows that this disadvantage precludes a successful invasion by asexuals with a phenotype that is optimal for region number 5, or by asexuals with phenotypes optimal for any of the regions south of region 5. However, a successful invasion is possible by an asexual phenotype that is optimal for any of the regions north of region 5. This pattern of asexuals being able to invade the initial equilibrium successfully only if they are adapted to northern regions was often found among the polymorphic trials (about 25% of the time). However, we also found that, in about 75% of the polymorphic trials, asexuals optimal for any region could invade the initial equilibrium successfully. Nevertheless, even among these trials, the probability of establishment of a clone after its introduction was generally higher if the clone was adapted to the north and was introduced in the north than if the clone was adapted to the south and introduced in the south.

What accounts for the difference in the likely fate of north- and south-adapted asexual clones? Roughly speaking, members of these clones do not mate, and so individuals with the optimal phenotype for a given region produce offspring with this same optimal phenotype. In contrast, a sexual adult with the locally optimal phenotype may have a non-optimal mate (possibly a migrant), and so many of the offspring may also be non-optimal. Immigrants and their recent descendants are most common in the north, so the initial invasion of asexuals is more likely in northern areas (where their mode of reproduction confers the largest advantage) than in the south. A similar explanation has been proposed to account for the presence of self-fertilizing plants in areas with unusual local environments²¹. Note that a loss of fitness due to maladapted migrants is a well-known phenomenon in natural populations^{21,22}, and has recently been implicated in speciation and the establishment of the limits of species ranges^{23,24}.

Throughout the series of invasions that led to the equilibrium shown in Fig. 1a, asexual clones adapted to the far south were at a disadvantage compared with some north-adapted clones (see Fig. 2). The reasons for this are similar to those state above with regard to the initial all-sexual equilibrium.

But what accounts for the 77 polymorphic trials (4%) in which the average position for asexuals was further south than that for sexuals at the final equilibrium? The answer appears to have something to do with the occurrence of unlikely events in the series of invasions leading to the final equilibrium. Strong support for this idea comes from tests in which we re-ran each of the exceptional trials 10 times. In all but 14 of the 77 cases, the average position of sexuals was further south than that of asexuals at the final equilibrium for most of the 10 trials run.

In addition to the simulation study just described, we have also carried out a set of simulations using a model in which the phenotype of the clone that makes the first successful invasion is as described above, but all the subsequent clones are produced as mutations to randomly chosen sexual individuals (just like the first clone), not as mutations to asexual individuals. The results were similar to those described above.

As we have seen, the tendency of the model to produce patterns similar to those found in nature depends on the presence of relatively short growing seasons in the north, and this leads to

high frequencies of migrants in northern areas. However, short growing seasons also occur at high altitudes. Thus the processes described here may also explain the tendency of asexuals to occur at high altitudes. Migrants should also be common on the margins of populations because the population density in these areas is often very low compared with nearby (but more central) regions. It has often been asserted that asexuals tend to occur on the margins of populations²⁻⁸, and given our results it seems likely that this is a result of a process similar to the one described above. We have now carried out a simulation study that supports this view (results not shown). □

Received 8 September; accepted 5 November 1997.

- Vandel, A. La parthénogénèse géographique: contribution l'étude biologique et cytologique de la parthénogénèse naturelle. *Bull. Biol. Belg.* **62**, 164–281 (1928).
- Suomalainen, E. Parthenogenesis in animals. *Adv. Genet.* **3**, 193–253 (1950).
- Lynch, M. Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. *Q. Rev. Biol.* **59**, 257–290 (1984).
- Bierzychudek, P. Patterns in plant parthenogenesis. *Experientia* **41**, 1255–1263 (1985).
- Hughes, R. N. *A Functional Biology of Clonal Animals* (Chapman and Hall, London, 1989).
- Glesener, R. R. & Tilman, D. Sexuality and the components of environmental uncertainty: Clues from geographic parthenogenesis in terrestrial animals. *Am. Nat.* **112**, 659–673 (1978).
- Bell, G. *The Masterpiece of Nature* (Univ. California Press, San Francisco, 1982).
- Parker, E. D. in *Evolutionary Genetics from Molecules to Morphology* (ed. Singh, R. & Krimbas, C.) (Cambridge Univ. Press, in the press).
- Bulmer, M. *The Mathematical Theory of Quantitative Genetics* (Clarendon, Oxford, 1980).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
- Charlesworth, B. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
- Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl Acad. Sci. USA* **87**, 3566–3573 (1990).
- Hurst, L. D. & Peck, J. R. Recent advances in understanding of the evolution and maintenance of sex. *Trends Ecol. Evol.* **11**, 46–52 (1996).
- Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
- Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**, 372–387 (1993).
- Peck, J. R. Frequency dependent selection, beneficial mutations, and the evolution of sex. *Proc. R. Soc. Lond. B* **125**, 87–92 (1993).
- Peck, J. R. A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**, 597–606 (1994).
- Peck, J. R. Limited dispersal, deleterious mutations and the evolution of sex. *Genetics* **142**, 1053–1060 (1996).
- Peck, J. R., Barreau, G. & Heath, S. C. Imperfect genes. Fisherian mutation and the evolution of sex. *Genetics* **145**, 1171–1199 (1997).
- Michod, R. E. & Levin, B. R. *The Evolution of Sex: an Examination of Current Ideas* (Sinauer, Sunderland, MA, 1988).
- Antonovics, J. Evolution in closely adjacent plant populations. V. Evolution of self-fertility. *Heredity* **23**, 219–238 (1968).
- Dias, P. C. Sources and sinks in population biology. *Trends Ecol. Evol.* **11**, 326–330 (1996).
- Kirkpatrick, M. & Barton, N. H. Evolution of a species range. *Am. Nat.* (in the press). (Author: update?)
- García-Ramos, G. & Kirkpatrick, M. Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**, 21–28 (1997).
- Barton, N. H. The probability of establishment of an advantageous mutant in a subdivided population. *Genet. Res.* **50**, 35–40 (1987).

Acknowledgements. We thank N. H. Barton, B. Charlesworth, J. Maynard Smith, D. Parker and M. Turelli for advice and assistance.

Correspondence and requests for materials should be addressed to J.R.P. at School of Biological Sciences, University of Sussex, Brighton BN1 9QG.

Activity-dependent scaling of quantal amplitude in neocortical neurons

Gina G. Turrigiano, Kenneth R. Leslie, Niraj S. Desai, Lana C. Rutherford & Sacha B. Nelson

Department of Biology and Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02254, USA

Information is stored in neural circuits through long-lasting changes in synaptic strengths^{1,2}. Most studies of information storage have focused on mechanisms such as long-term potentiation and depression (LTP and LTD), in which synaptic strengths change in a synapse-specific manner^{3,4}. In contrast, little attention has been paid to mechanisms that regulate the total synaptic

strength of a neuron. Here we describe a new form of synaptic plasticity that increases or decreases the strength of all of a neuron's synaptic inputs as a function of activity. Chronic blockade of cortical culture activity increased the amplitude of miniature excitatory postsynaptic currents (mEPSCs) without changing their kinetics. Conversely, blocking GABA (γ -aminobutyric acid)-mediated inhibition initially raised firing rates, but over a 48-hour period mEPSC amplitudes decreased and firing rates returned to close to control values. These changes were at least partly due to postsynaptic alterations in the response to glutamate, and apparently affected each synapse in proportion to its initial strength. Such 'synaptic scaling' may help to ensure that firing rates do not become saturated during developmental changes in the number and strength of synaptic inputs⁵, as well as stabilizing synaptic strengths during Hebbian modification^{6,7} and facilitating competition between synapses⁷⁻⁹.

We tested the idea that postnatal rat visual cortical pyramidal neurons in primary cell culture scale the strength of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-mediated synaptic currents up or down as a function of activity. To sample inputs from a large number of synapses we used whole-cell voltage-clamp recordings to measure mEPSCs. In control cultures there was a broad distribution of mEPSC amplitudes^{10,11} (Fig. 1a; average quantal amplitude = 13.6 ± 0.7 pA at -70 mV, $n = 27$). The effects of decreasing or increasing firing rates for 48 h on mEPSC amplitude are shown in Fig. 1a. On average, mEPSC amplitudes from cultures grown in tetrodotoxin (TTX), which

completely abolished firing, increased to $192 \pm 16\%$ of control values, and there was a similar effect on area (Fig. 1b, $n = 10$). Blockade of GABA_A-mediated inhibition with bicuculline, which raised firing rates to $265 \pm 10\%$ of control values, significantly decreased mEPSC amplitude to $70 \pm 4\%$ of control values (Fig. 1b, $n = 10$), and also decreased mEPSC area. For the manipulations described here and below, resting potentials, input resistances, whole-cell capacitances and series resistances were similar between conditions (Table 1). These data demonstrate that the quantal amplitude of AMPA synapses can be increased or decreased by changes in activity. Activity blockade did not significantly affect the amplitude of AMPA-mediated mEPSCs recorded from bipolar interneurons (Fig. 1c, $n = 14$), indicating that excitatory inputs from pyramidal neurons are regulated differently depending on the nature of the target neuron.

Some forms of synaptic plasticity require calcium influx through NMDA (*N*-methyl-D-aspartate) receptors^{3,4}. To examine the effects of NMDA receptor blockade on mEPSC amplitude, cultures were grown in the NMDA antagonist AP5 (*D*(-)-amino-7-phosphonovalenic acid) for 48 h. AP5 did not influence firing rates ($95 \pm 24\%$ of control values), and had no effect on mEPSC amplitude or area (Fig. 1b, $n = 8$). In contrast, blocking AMPA receptors with 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX), which like TTX completely abolished firing, significantly increased mEPSC amplitude to $195 \pm 24\%$ of control values (Fig. 1b, $n = 10$). These observations indicate that, in contrast to many forms of LTP and LTD^{3,4,12}, the change in synaptic strength produced by activity blockade is not simply the result of a change in NMDA receptor signalling.

Cumulative amplitude histograms showed that the entire distribution of mEPSC amplitudes shifted towards larger values for

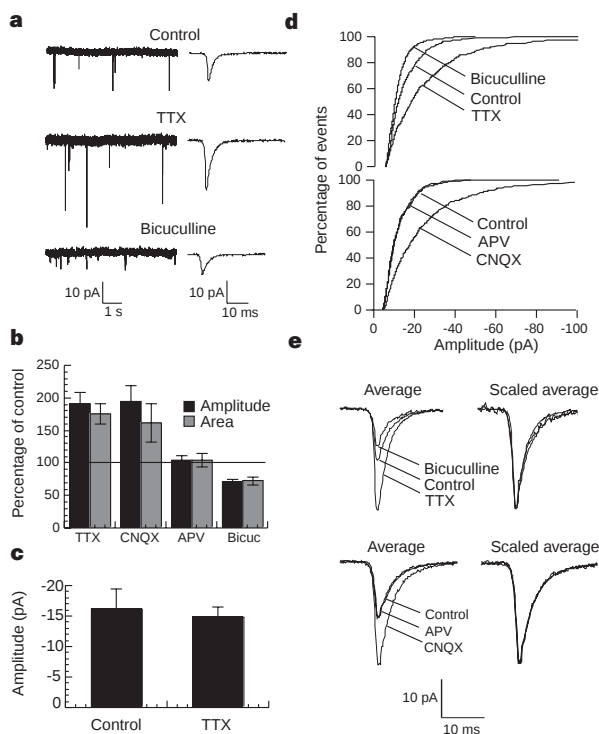


Figure 1 AMPA mEPSC amplitudes from control cultures or cultures grown in TTX, CNQX or bicuculline for 48 hours. **a**, Representative recordings. Left, raw data; right, average mEPSC waveform for the neuron on the left. **b**, Average mEPSC amplitudes and areas expressed as a percentage of control (one way analysis of variance (ANOVA), $P < 0.002$; *t*-tests (using a Bonferroni correction for multiple comparisons) indicate that TTX, CNQX and bicuculline (Bicuc) amplitudes differ significantly from control ($P < 0.004$, 0.02 and 0.01 , respectively), whereas AP5 does not. **c**, Average amplitudes of AMPA mEPSCs from bipolar interneurons grown with or without TTX for 48 h. **d**, Cumulative amplitude histograms for mEPSCs recorded under each condition. **e**, Average mEPSC waveforms for each condition, unscaled (average) and scaled (scaled average).

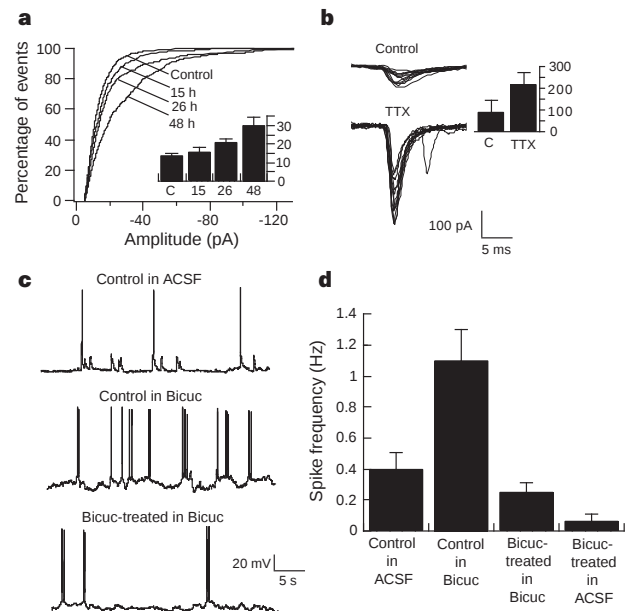


Figure 2 Effects of activity on mEPSCs, paired transmission, and firing rates. **a**, Time course of changes in quantal amplitude. Inset, mean mEPSC amplitudes for control and after 15, 26 and 48 h (one way ANOVA, $P < 0.002$). **b**, Paired recordings from control and TTX-treated cultures. Inset, mean $\pm$ s.e.m. for control ($n = 13$) or TTX-treated ($n = 9$, $P < 0.005$, Wilcoxon rank sum test) pairs. **c**, **d**, Regulation of firing rates during chronic exposure to bicuculline. Control cultures recorded in ACSF or bicuculline (Bicuc), and bicuculline-treated cultures recorded in bicuculline or ACSF. **c**, Representative recordings. **d**, mean $\pm$ s.e.m. ($n = 9$ in each condition; Bicuc in Bicuc significantly different from control in Bicuc, $P < 0.01$, and not significantly different from control in ACSF, $P = 0.24$, corrected *t*-test).

TTX-treated and CNQX-treated cultures, and towards smaller values for bicuculline-treated cultures (Fig. 1d). We examined mEPSC kinetics by determining the average mEPSC waveforms under the different conditions (Fig. 1e). Scaling and overlaying these averaged waveforms revealed no differences in the rise and decay kinetics (Fig. 1e), and individual measurements of rise times also showed no significant differences between conditions (Table 1). In contrast to LTP and other rapid potentiation protocols^{13–15}, there were no significant differences in mEPSC frequency between conditions (Table 1). The lack of effect on frequency and kinetics suggests that, in accord with other studies^{16,17}, altering activity for two days does not significantly influence the placement or number of excitatory synapses onto pyramidal neurons.

The time course of mEPSC regulation was determined by blocking activity for 15, 26, or 48 (±4) h (*n* = 8, 11 and 10, respectively). There was a progressive increase in mEPSC amplitude with treatment time, indicating that this process is both slow and cumulative (Fig. 2a).

Does the activity-dependent regulation of mEPSC amplitude produce a change in spike-mediated EPSCs? TTX treatment increased

EPSC amplitude between pairs of monosynaptically connected pyramidal neurons from 89 ± 65 pA (measured at -70 mV, *n* = 13 pairs) to 219 ± 52 pA (*n* = 9 pairs) (Fig. 2b). There was no effect on EPSC reversal potential (1.6 ± 1.0 and 2.2 ± 1.1 mV for control and TTX-treated pyramidal neurons, respectively).

The regulation of mEPSC amplitudes described above could act to stabilize firing rates, because raising firing rates decreases the strength of excitatory synaptic inputs, and vice versa. To test whether firing rates are regulated homeostatically in response to changes in activity, cultures were grown under control conditions or in bicuculline for 48 h, and whole-cell current-clamp recordings were used to measure firing rates (Fig. 2c). Acute bicuculline treatment raised firing rates significantly from 0.4 ± 0.1 to 1.1 ± 0.2 Hz (Fig. 2d), but after 48 h in bicuculline, firing rates recorded in bicuculline declined significantly to 0.24 ± 0.06 Hz, and were close to zero when bicuculline was removed (Fig. 2d). These data indicate that increased firing produces a regulatory response that returns firing rates to control levels. Previous studies have shown that chronic TTX treatment dramatically increases firing rates when the TTX is removed^{18,19}. The bidirectional regulation of

Table 1 Effects of treatments on neuronal properties

Treatment	<i>V_m</i>	<i>R_{in}</i>	Rise time	Capacitance	<i>S_r</i>	Frequency
TTX	99.8 ± 3.5	113.0 ± 19.0	91.6 ± 8.4	104.1 ± 11.7	101.1 ± 9.7	95.5 ± 15.5
CNQX	105.5 ± 4.2	108.4 ± 18.2	92.8 ± 6.4	95.5 ± 9.2	83.6 ± 8.9	99.2 ± 26.0
Bicuculline	103.5 ± 2.3	100.3 ± 14.0	103.5 ± 2.3	99.8 ± 8.6	110.4 ± 6.6	108.2 ± 22.3
AP5	102.3 ± 2.6	102.3 ± 5.0	91.6 ± 9.7	104.7 ± 11.7	100.2 ± 11.2	99.5 ± 26.5

Effects of treatment on resting potential (*V_m*), input resistance (*R_{in}*), mEPSC rise times, whole-cell capacitance, series resistance (*S_r*) and mEPSC frequency, expressed as a percentage of values from sister control cultures. None of these differences were statistically significant. For controls, *V_m* = 60.3 ± 1.6 mV; *R_{in}* = 1.1 ± 0.4 GΩ; rise time = 0.78 ± 0.05 ms; capacitance = 28.7 ± 3.9 pF; *S_r* = 12.8 ± 1.2 MΩ; and mEPSC frequency = 0.34 ± 0.08 Hz.

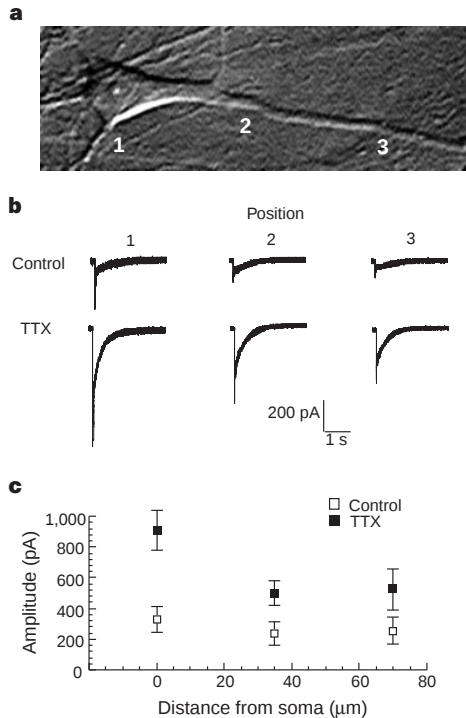


Figure 3 Activity blockade increases the postsynaptic sensitivity to glutamate. **a**, Differential interference contrast image of a cultured pyramidal neuron, showing the sites at which glutamate was applied: 1, soma; 2, 35 μm from the soma; and 3, 70 μm from the soma. **b**, Representative responses for control and TTX-treated neurons: 1, 2 and 3 correspond to the sites indicated in **a**. Each trace shows 5 overlaid responses from the same site. **c**, Average responses at the three sites for control (*n* = 9) and TTX-treated (*n* = 8) neurons. Responses for TTX-treated neurons are significantly larger than control (site 1, *P* < 0.001; 2, *P* < 0.02; 3, *P* < 0.05, *t*-test).

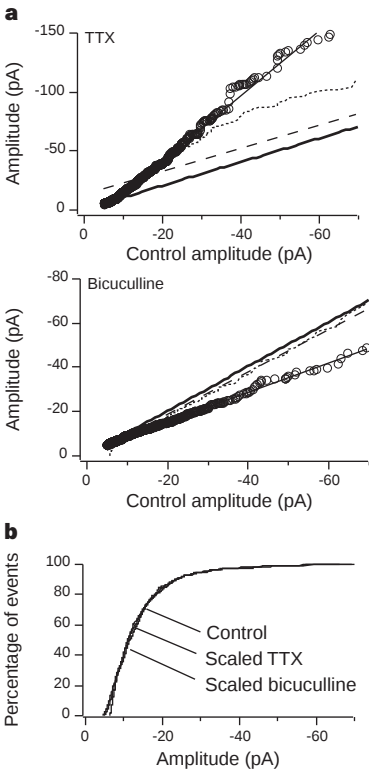


Figure 4 Activity scales mEPSC amplitudes multiplicatively. **a**, Ranked control amplitudes were plotted against ranked TTX or bicuculline amplitudes, and the best fit to the data (open circles) of additive (dashed lines), random additive (dotted curves), or multiplicative (thin solid lines) functions was determined. Thick solid lines are control plotted against control, and have slopes of 1. Best fit: TTX = control × 2.73 + 12.3, *R* = 0.997; Bicuculline = control × 0.66 - 2.0, *R* = 0.998. **b**, The original TTX and bicuculline distributions were transformed by the preceding equations and plotted with the control data.

quantal amplitude is likely to contribute significantly to the homeostatic regulation of firing rates.

Changes in mEPSC amplitude could be occurring through a postsynaptic change in glutamate responsiveness or a presynaptic change in the glutamate content of synaptic vesicles. To investigate this, pulses of glutamate were applied to pyramidal neurons in the presence of TTX and AP5 and the amplitudes of the resulting glutamate currents were measured (Fig. 3a, b). Chronic TTX treatment increased the current amplitude to $278 \pm 39\%$ of control values at the soma, and similar increases were seen along the apical dendrite (Fig. 3c; control, $n = 9$; TTX, $n = 8$). No difference in rise times or times to peak of the glutamate currents were observed between conditions, indicating that the difference in amplitude is not due to a difference in rates of diffusion or removal of glutamate under the two conditions (rise times = 7.5 ± 0.7 and 8.9 ± 0.8 ms, and times to peak = 23 ± 1.4 and 25 ± 1.1 ms, respectively). These data suggest that activity regulates mEPSC amplitude through a postsynaptic change in receptor number or function, although there may be additional presynaptic changes. In contrast to supersensitivity at the neuromuscular junction, where enhanced responsiveness is exclusively extrajunctional and mEPP amplitude does not change^{20,21}, the change in mEPSC amplitude observed here indicates that receptors located at the synapse are involved in the regulatory process.

What is the relationship between the mEPSC amplitude distributions produced by different activity levels? The average quantal amplitude could be modified by changing synaptic strengths by a constant value (an additive function), by adding a random amount (a random additive function), or by scaling synaptic strengths by the same multiplicative factor (a multiplicative function). To distinguish between these possibilities, control amplitudes were plotted against TTX amplitudes (Fig. 4a) or bicuculline amplitudes (Fig. 4b) and the resulting relationship fit using each of the above models. The data were well fit by linear functions with slopes of 2.73 (TTX) and 0.66 (bicuculline), but not by additive or random additive functions (Fig. 4a). When scaled multiplicatively by these slope values, the cumulative amplitude distributions for TTX and bicuculline were almost perfectly superimposable on the control distribution (Fig. 4b), suggesting that activity is globally scaling quantal amplitudes in a multiplicative manner. Taken together, our data place several constraints on the biophysical mechanism of synaptic scaling. If this scaling occurs through a change in the function of existing AMPA receptors through phosphorylation^{22,23} or some other mechanism, this process must modify single-channel conductance without influencing mEPSC kinetics. Alternatively, if this scaling occurs through changes in the synthesis and insertion of glutamate receptors^{24,25}, or through conversion of existing receptors between inactive and active states^{26,27}, then receptors must be inserted or converted in proportion to the existing number of functional receptors.

Multiplicative scaling of synaptic strengths preserves relative differences between inputs (such as those produced by Hebbian modifications), while allowing a neuron to adjust the total amount of synaptic excitation it receives. This process may help developing neurons remain responsive to inputs both early in development, when the number of synaptic inputs is small, and later in development as the number rises²⁸. This may be important for allowing developing neurons to participate in the correlation-based mechanisms that contribute to circuit formation⁵. Consistent with this, there is evidence that mEPSC amplitudes in hippocampal neurons decrease as the number of synaptic inputs increase²⁹. In addition, by contributing to a stabilization in firing rates, synaptic scaling may help to counteract the destabilizing effects of Hebbian synaptic modifications^{6,7,30}. This instability arises because synaptic potentiation increases postsynaptic firing rates, thus increasing the correlation with presynaptic activity and producing a positive feedback loop that eventually saturates even weakly correlated inputs. Finally,

this process predicts that strengthening of some inputs will lead to a scaling down of all synaptic strengths. Synaptic weakening has been suggested to be the prelude to synapse elimination^{8,9}. By reducing weak inputs in response to the strengthening of others, synaptic scaling may contribute to the processes of synaptic competition and elimination. □

Methods

Rat visual cortical cultures were prepared from rat pups at postnatal days 4–6, and whole-cell recordings were obtained in artificial cerebrospinal fluid (ACSF) as described¹⁸. Experiments were performed after 7–9 days *in vitro*. All data were obtained in parallel on treated and age-matched sister control cultures. Drug concentrations were 1 μ M TTX, 20 μ M bicuculline, 20 μ M CNQX and 50 μ M AP5. Recordings with resting potentials greater than -55 mV, series resistances greater than 20 M Ω , or fewer than 50 mEPSCs were excluded. Pyramidal and non-pyramidal neurons were distinguished as described¹⁸. To record mEPSCs, neurons were held in voltage clamp at -70 mV using an Axopatch 1D in the presence of TTX, bicuculline and AP5. In-house software was used to detect and measure mEPSCs; detection criteria included amplitudes greater than 5 pA and 20–80% rise times less than 3 ms. All data are reported as mean $\pm$ s.e.m. for the number of neurons indicated. Paired recordings were obtained in ACSF with AP5, and the pipette solution for the postsynaptic cell contained caesium and QX-314. Monosynaptic EPSCs were considered to be those with latencies of less than 4 ms, less than 1 ms of jitter, a monotonic rising phase, and reversal potentials of approximately 0 mV. A picrospritzer was used to deliver puffs of 0.5 mM glutamate (pipette diameter, 2 μ m) while holding pyramidal neurons at -70 mV in whole-cell voltage clamp in the presence of TTX and AP5. Puffs were repeated 5 times at each site and the results averaged. The relationship between control and experimental amplitude distributions was determined by ranking mEPSC amplitudes from all cells recorded from TTX-treated, bicuculline-treated, or control cultures in ascending order. These amplitude distributions were linearly interpolated to produce an equal number of observations in each condition, and experimental amplitudes plotted against control amplitudes. For additive function, the same value was added to each control amplitude; for random additive functions, the differences between ranked experimental and control amplitudes were calculated and these values were added to control amplitudes in random order; for multiplicative functions, each control value was transformed by a linear equation with a slope greater than (for TTX) or less than (bicuculline) 1; the non-zero intercepts of the best-fit multiplicative functions are close to those required (8.7 and -1.7 pA, respectively) to compensate for the 5-pA amplitude cut-offs.

Received 5 November 1997; accepted 19 January 1998.

- Hebb, D. O. *The Organization of Behavior: a Neurophysiological Theory* (Wiley, New York, 1949).
- Stent, G. S. A physiological mechanism for Hebb's postulate of learning. *Proc. Natl Acad. Sci. USA* **70**, 997–1001 (1973).
- Madison, D. V. *et al.* Mechanisms underlying long-term potentiation of synaptic transmission. *Annu. Rev. Neurosci.* **14**, 379–397 (1991).
- Linden, D. J. & Connor, J. A. Long-term synaptic depression. *Annu. Rev. Neurosci.* **18**, 319–357 (1995).
- Shatz, C. J. Impulse activity and the patterning of connections during CNS development. *Neuron* **5**, 745–756 (1990).
- Bienenstock, E. L. *et al.* Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
- Miller, K. D. Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron* **17**, 371–374 (1996).
- Lo, Y. J. & Poo, M. Activity-dependent synaptic competition in vitro: heterosynaptic suppression of developing synapses. *Science* **254**, 1019–1022 (1991).
- Colman, H. *et al.* Alterations in synaptic strength preceding axon withdrawal. *Science* **275**, 356–361 (1997).
- Bekkers, J. M. *et al.* Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl Acad. Sci. USA* **87**, 5359–5362 (1990).
- Hestrin, S. Activation and desensitization of glutamate-activated channels mediating fast excitatory synaptic currents in the visual cortex. *Neuron* **9**, 991–999 (1992).
- Kirkwood, A. & Bear, M. F. Hebbian synapses in visual cortex. *J. Neurosci.* **14**, 1634–1645 (1994).
- Manabe, T. *et al.* Postsynaptic contribution to long-term potentiation revealed by the analysis of miniature synaptic currents. *Nature* **355**, 50–55 (1992).
- Malgaroli, A. *et al.* Glutamate-induced long-term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature* **357**, 134–139 (1992).
- Wyllie, D. J. *et al.* A rise in postsynaptic calcium potentiates miniature excitatory postsynaptic currents and AMPA responses in hippocampal neurons. *Neuron* **12**, 127–138 (1994).
- Micheva, K. D. & Beaulieu, C. An anatomical substrate for experience-dependent plasticity of the rat barrel field cortex. *Proc. Natl Acad. Sci. USA* **92**, 11834–11838 (1995).
- Benson, D. L. & Cohen, P. A. Activity-independent segregation of excitatory and inhibitory synaptic terminals in cultured hippocampal neurons. *J. Neurosci.* **16**, 6424–6432 (1996).
- Rutherford, L. C. *et al.* BDNF mediates the activity-dependent regulation of inhibition in neocortical cultures. *J. Neurosci.* **17**, 4527–4535 (1997).

19. Ramakers, G. J. A. *et al.* Development in the absence of spontaneous bioelectric activity results in increased stereotyped burst firing in cultures of dissociated cerebral cortex. *Exp. Brain Res.* **79**, 157–166 (1990).
20. Axelsson, J. & Thesleff, S. A study of supersensitivity in denervated mammalian skeletal muscle. *J. Physiol. (Lond.)* **149**, 178–193 (1957).
21. Berg, D. K. & Hall, Z. W. Increased extrajunctional acetylcholine sensitivity produced by chronic post-synaptic neuromuscular blockade. *J. Physiol. (Lond.)* **244**, 659–676 (1975).
22. McGlade-McCulloh, E. *et al.* Phosphorylation and regulation of glutamate receptors by calcium/calmodulin-dependent protein kinase II. *Nature* **362**, 640–642 (1993).
23. Barria, A. *et al.* Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science* **276**, 2042–2045 (1997).
24. Gerfin-Moser, A. *et al.* Alterations in glutamate but not GABA_A receptor subunit expression as a consequence of epileptiform activity in vitro. *Neuroscience* **67**, 849–865 (1995).
25. Liang, F. & Jones, E. G. Differential and time-dependent changes in gene expression for type II calcium/calmodulin-dependent protein kinase, 67 kDa glutamic acid decarboxylase, and glutamate receptor subunits in tetanus toxin-induced focal epilepsy. *J. Neurosci.* **17**, 2168–2180 (1997).
26. Liao, D. *et al.* Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
27. Wu, G.-Y. *et al.* Maturation of a central glutamatergic synapse. *Science* **274**, 972–976 (1996).
28. Blue, M. E. & Parnavelas, J. G. The formation and maturation of synapses in the visual cortex of the rat. II. Quantitative analysis. *J. Neurocytol.* **12**, 697–712 (1983).
29. Liu, G. & Tsien, R. W. Properties of synaptic transmission at single hippocampal synaptic boutons. *Nature* **375**, 404–408 (1995).
30. Miller, K. D. & MacKay, D. J. C. The role of constraints in Hebbian learning. *Neur. Comput.* **6**, 100–126 (1993).

Acknowledgements. We thank H. Lauer for technical assistance, and L. F. Abbott, J. Lisman, E. Marder and M. Mauk for discussions and advice. This work was supported by the Whitehall Foundation, NSF, NIH and the Sloan Foundation. K.R.L. was supported by an HHMI predoctoral fellowship.

Correspondence and requests for materials should be addressed to G.G.T. (e-mail: turrigiano@binah.cc.brandeis.edu).

Molecular characterization of a neuronal low-voltage-activated T-type calcium channel

Edward Perez-Reyes^{*†}, Leanne L. Cribbs^{*†}, Asif Daud^{*}, Antonio E. Lacerda[‡], Jane Barclay[§], Magali P. Williamson[§], Margaret Foxll, Michele Rees[§] & Jung-Ha Lee^{*}

^{*} Department of Physiology, and [†] Cardiovascular Institute, Loyola University Medical Center, Maywood, Illinois 60153, USA

[‡] Rammelkamp Center for Research & Education, MetroHealth Medical Center, Cleveland, Ohio 44109, USA

[§] Department of Paediatrics, The Rayne Institute, University College London Medical School, London WC1E 6JJ, UK

ll MRC Human Biochemical Genetics Unit, The Galton Laboratory, London NW1 2HE, UK

The molecular diversity of voltage-activated calcium channels was established by studies showing that channels could be distinguished by their voltage-dependence, deactivation and single-channel conductance^{1–3}. Low-voltage-activated channels are called 'T' type because their currents are both transient (owing to fast inactivation) and tiny (owing to small conductance)². T-type channels are thought to be involved in pacemaker activity, low-threshold calcium spikes, neuronal oscillations and resonance, and rebound burst firing⁴. Here we report the identification of a neuronal T-type channel. Our cloning strategy began with an analysis of Genbank sequences defined as sharing homology with calcium channels. We sequenced an expressed sequence tag (EST), then used it to clone a full-length complementary DNA from rat brain. Northern blot analysis indicated that this gene is expressed predominantly in brain, in particular the amygdala, cerebellum and thalamus. We mapped the human gene to chromosome 17q22, and the mouse gene to chromosome 11. Functional expression of the channel was measured in *Xenopus* oocytes. Based on the channel's distinctive voltage dependence, slow deactivation kinetics, and 7.5-pS single-channel conductance, we conclude that this channel is a low-voltage-activated

T-type calcium channel.

The molecular biology of voltage-activated Ca²⁺ channels began with the cloning of the skeletal muscle dihydropyridine receptor, $\alpha 1S$ (reviewed in ref. 5). Using a combination of low-stringency hybridization and polymerase chain reaction (PCR) techniques, five other $\alpha 1$ subunits of calcium channels have been cloned⁵. All six of these $\alpha 1$ subunits form high-voltage-activated (HVA) Ca²⁺ channels. Original studies on a rat $\alpha 1E$ clone⁶ suggested that it belonged to the low-voltage-activated (LVA) family. However, mouse⁷, rabbit⁸ and human^{7,9} $\alpha 1E$ clones form HVA channels. Strong evidence that $\alpha 1E$ is not part of a T-type channel was provided by studies showing it had double the single channel conductance for Ba²⁺ (12–14 pS)^{7–10}. Our alternative cloning strategy was to use a text-based search of the Genbank to find new sequences that showed homology with cloned calcium channels. Based on its 45% homology with domain III S1 of the carp $\alpha 1$ (P22316), we decided to sequence H06096. Although its deduced amino-acid sequence had low homology to cloned Ca²⁺ channels, it contained readily identifiable motifs including an S4 region and a pore loop. Using H06096 as a probe, we screened a rat brain λ gt10 cDNA library. A full-length cDNA, referred to as $\alpha 1G$ or Ca_vT.1, was assembled from five overlapping clones. Search of the non-redundant division of the Genbank revealed that $\alpha 1G$ was similar to other Ca²⁺ channel $\alpha 1$ subunits, but most similar to C54D2.5 (Genbank No. U37548), a putative protein found in the genomic DNA of *Caenorhabditis elegans*¹¹. Homologous human (H19230, R19524) and mouse (AA386626) EST clones were also identified and sequenced. Sequence identity among the Ca²⁺ channel $\alpha 1$ subunits is highest in the putative membrane-spanning regions, with most changes being conservative with respect to structure (Fig. 1). Charged residues are particularly conserved, with many charges being conserved across all domains and in voltage-gated Na⁺ and K⁺ channels¹². The charged residues of the S4 regions are also conserved, consistent with its role as a voltage sensor¹³. The cation selectivity of Ca²⁺ channels requires a ring of negative charge provided by glutamate residues found at similar locations in each domain¹⁴. In $\alpha 1G$, two of these glutamates are replaced by aspartate, suggesting an altered selectivity. In contrast, there is little conservation of the sequences that link these regions within a domain, and even less between the intracellular loops that connect the domains. Notably, the motifs involved in binding the β -subunit^{15,16} and calcium¹⁷ are missing.

Northern blot analysis of human and rat tissues indicates there are two $\alpha 1G$ messenger RNA transcripts, with a predominant band of 8.5 kilobases (kb) and another of 9.7 kb (Fig. 2). The strongest signals were detected in brain, and less abundantly in heart. Longer exposure reveals expression in placenta, kidney and lung. Transcripts were detected in all human brain regions studied, with strong signals from (in relative order of abundance) amygdala, thalamus, subthalamic nuclei (Fig. 2c) and cerebellum (data not shown). To confirm the northern blot analysis, we used PCR followed by subcloning and sequencing.

Genetic mutations of either the calcium channel $\alpha 1A$ subunit or β_4 subunits result in ataxic and epileptic phenotypes^{18,19}. A similar link between T-type channels and absence epilepsy has been suggested²⁰. As a first step in exploring such a link we mapped the human and murine $\alpha 1G$ genes. The human $\alpha 1G$ gene, CACNA1G, was mapped to chromosome 17 between the markers AFMA126YD5 and D17S798 (LOD score > 3.0), using the Genbridge 4 radiation hybrid panel. Fluorescent *in situ* hybridization to normal male human metaphase spreads was carried out using the $\alpha 1G$ cDNA clone R19524. Analysis of 22 metaphases confirmed the localization of CACNA1G to chromosome 17 and 59% of these could be assigned to the 17q22 band. The mouse $\alpha 1G$ gene, *Cacna1g*, was mapped to chromosome 11 on the EUCIB map²¹ between the markers JPAV507 and D11Mit263 with a LOD score of 8.0.

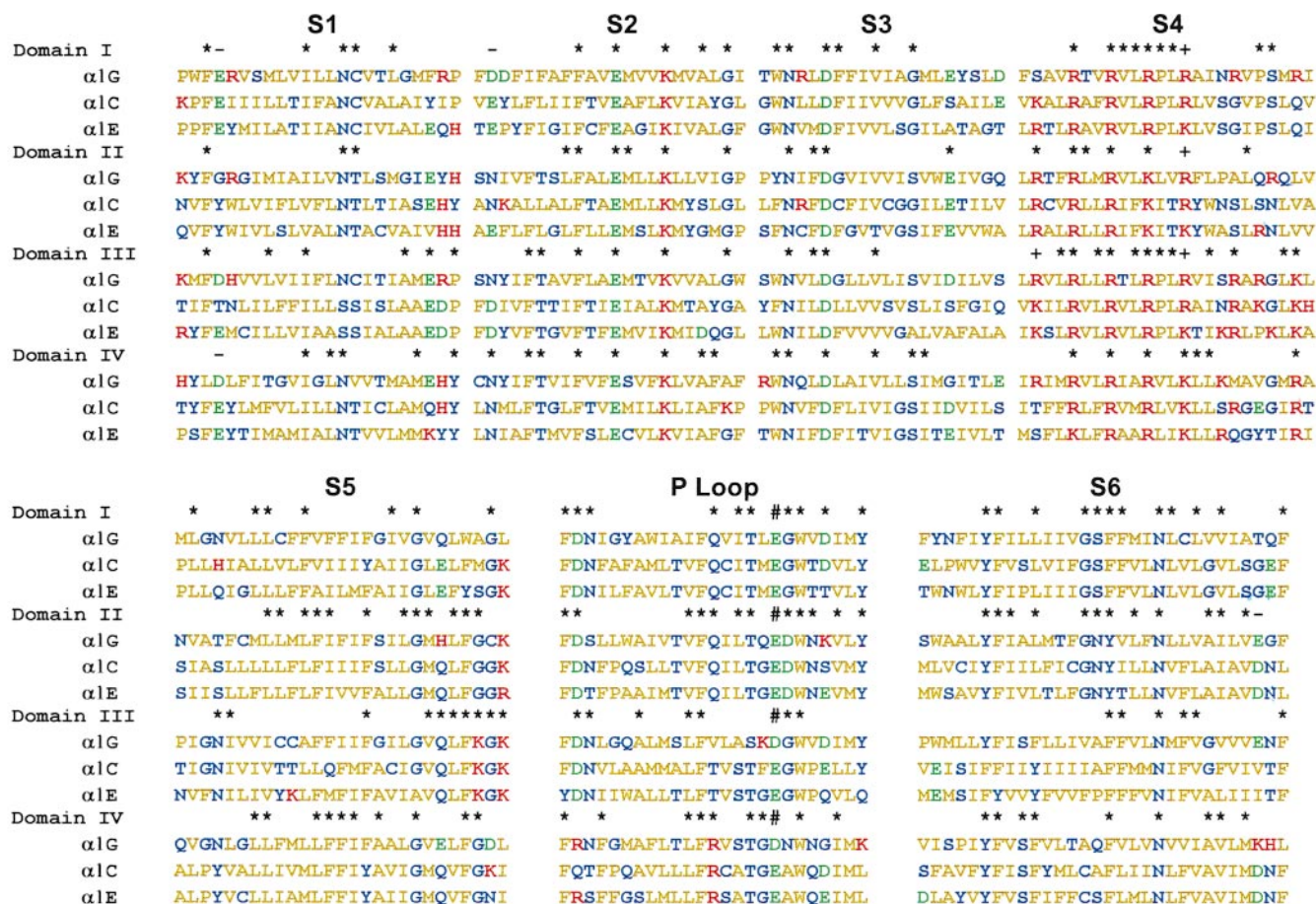


Figure 1 Alignment of the putative membrane spanning regions of α1G, α1C and α1E. Amino acids are colour coded as follows: positively charged residues are red; negatively charged residues are green; neutral, polar residues are blue; and neutral, non-polar residues are yellow. Amino acids that are conserved in all seven mammalian α1 subunits of voltage-activated Ca²⁺ channels are marked with an asterisk. Charge conservation is marked with a plus or minus sign. The glutamate residues in the pore loops that determine Ca²⁺ selectivity in HVA Ca²⁺ channels¹⁴ are marked with a hash symbol. The GenBank accession numbers of the human sequences were: α1C, L04569; α1E, L27745; α1S, L33798; α1D, M76558; α1A, X99897; and α1B, M94172.

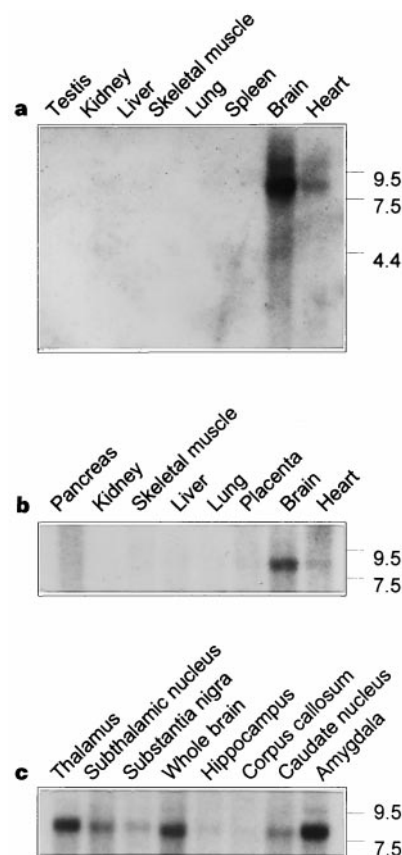


Figure 2 Distribution of α1G mRNA in rat and human. **a**, A rat multiple tissue northern blot was probed with α1G (nt 5,782–7,149), then exposed for 24 h. **b**, A human multiple tissue northern blot was probed with human H06096, which is 89% identical to the rat α1G (nt 3,589–4,545), then exposed for one week. **c**, Human brain region northern blot was probed and exposed as in **b**. Molecular mass markers are indicated on the right in kilobases. At least two independent blots were hybridized with probes from both transmembrane and 3' untranslated regions of each probe, giving similar results, thereby minimizing the possibility that the signals are due to crossreactivity with another mRNA.

Functional expression of $\alpha 1G$ was measured in *Xenopus* oocytes. Figure 3a shows a representative family of current traces elicited by depolarizing pulses of the oocyte. Inward Ba^{2+} currents activated slowly near threshold potentials (-60 mV in 2 mM Ba^{2+}), whereas stronger depolarizations produced a current that activated and inactivated quickly (Fig. 3e). Such crossing of successive currents has been called a signature pattern of classical T-type channels²². To illustrate this pattern (Fig. 3b) we recorded genuine T-type currents from undifferentiated NIE-115 mouse neuroblastoma cells²³. In contrast, currents from oocytes injected with $\alpha 1E$ have a distinct pattern (Fig. 3c), because activation is uniformly fast. Uninjected oocytes had no detectable inward Ba^{2+} currents (<10 nA). Figure 3d shows the current-voltage curves for $\alpha 1G$, $\alpha 1E$ and the T-type currents from NIE-115 cells (10 mM Ba^{2+}). These data were trans-

formed into conductance, then fitted using the Boltzmann equation to calculate the mid-point of activation ($V_{0.5}$). Both NIE-115 and $\alpha 1G$ currents gated at low voltages (NIE-115, -41 ± 1 mV, $n = 10$; $\alpha 1G$, -38 ± 1 , $n = 8$), whereas $\alpha 1E$ required more positive potentials (-2.6 ± 0.4 , $n = 3$). To help with comparison of these results to published values⁴, we recorded $\alpha 1G$ currents using varying concentrations of charge carrier. In solutions containing 2 mM Ba^{2+} , $V_{0.5}$ was -46.5 mV, and the slope factor (k) was 6.6 ($n = 7$). We routinely used solutions containing 40 mM Ba^{2+} because the currents recorded from most $\alpha 1G$ -injected oocytes were <0.5 Amps/Farads. Because of the effect of barium on surface charge screening, $V_{0.5}$ was shifted to -21 mV in 40 mM Ba^{2+} . These results demonstrate that rat brain $\alpha 1G$ is the first low-voltage-activated channel to be cloned. In addition, $\alpha 1G$ currents activate

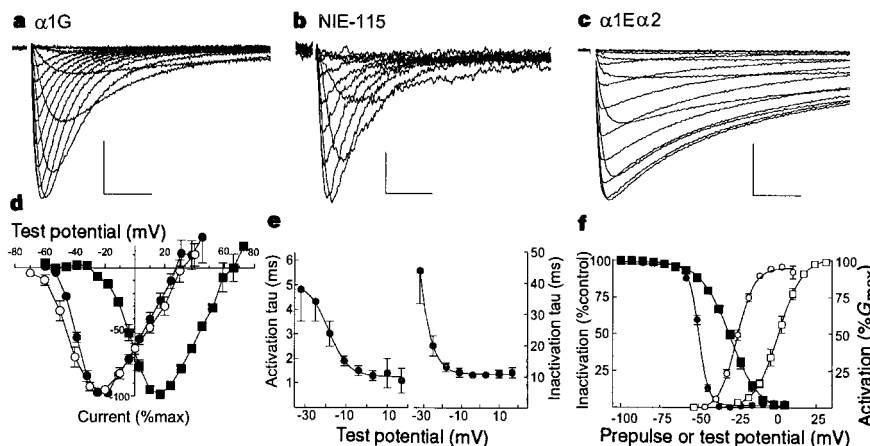


Figure 3 Current-voltage (I - V) relationships of cloned $\alpha 1G$ are compared to native T-type currents in NIE-115 cells and cloned $\alpha 1E$ channels. **a**, Cloned $\alpha 1G$ currents were recorded using 40 mM Ba^{2+} as charge carrier. **b**, Currents from an undifferentiated mouse neuroblastoma cell were recorded using the nystatin-perforated patch method (10 mM Ba^{2+}). **c**, Cloned $\alpha 1E$ plus $\alpha 2$ currents were recorded from an oocyte injected with 50 nl of 25 mM BAPTA (2 mM Ba^{2+}). **d**, Average I - V curves recorded in 10 mM Ba^{2+} . Peak currents for each cell were normalized to the maximum current observed, then averaged. Error bars represent the s.e.m. from NIE-115 cells (open circles), or oocytes injected with

either $\alpha 1G$ (filled circles) or $\alpha 1E$ (filled squares). **e**, Voltage dependence of $\alpha 1G$ activation and inactivation time constants. Data represent the mean $\pm$ s.d. from 13 oocytes. **f**, Voltage dependence of inactivation and activation measured under the same ionic conditions (40 Ba^{2+}). Data represent the mean $\pm$ s.e.m. from 5 ($\alpha 1G$, circles) or 6 ($\alpha 1E$, squares) oocytes. Chord conductance was calculated using the apparent reversal potentials, normalized to the peak conductance observed ($\% G_{max}$), averaged, then plotted against the test potential. Scale bars: **a**, 200 nA; **b**, 50 pA; **c**, 800 nA; and 20 ms.

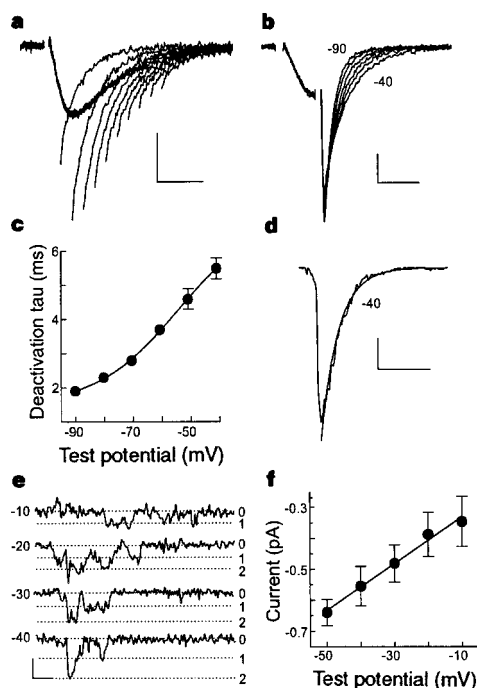


Figure 4 Deactivation and single-channel conductance of $\alpha 1G$ currents. **a**, Tail currents at -50 mV were measured after depolarizing pulses to -10 mV of varying duration. **b**, Voltage-dependence of deactivation was measured at varying potentials after a 5 ms depolarizing pulse to -10 mV. **c**, Mean data ($n = 3$) are plotted as a function of the repolarization potential. The voltage protocol included a 5 ms step to $+20$ mV followed by repolarization to the indicated potentials. **d**, The ensemble current was calculated from the idealized data from 100 sweeps, then fitted with a single exponential. **e**, Representative single channel openings at four potentials are shown. Numbers on the right indicate the number of channels open at any particular time. **f**, The current through the main open state of the channel was measured at each potential, then plotted versus the test potential. The slope of the line was calculated by linear regression. Scale bars: **a** and **b**, 500 nA, 5 ms; **d**, 0.25 pA, 25 ms; and **e**, 0.5 pA, 5 ms.

and inactivate with similar kinetics as observed for T-type channels (reviewed in ref. 4). These results were obtained after injection of $\alpha 1G$ alone, suggesting that other subunits are not required, or are supplied by the oocyte.

To determine the voltage dependence of steady-state inactivation, we measured channel availability after 10-s prepulses (Fig. 3f). Inactivation of $\alpha 1G$ occurred at sub-threshold potentials, and displayed a steep voltage dependence ($V_{0.5} = -50.0 \pm 0.2$ mV, $k = -3.2 \pm 0.2$, $n = 5$). Inactivation of $\alpha 1E$ occurred at more positive potentials and displayed a weaker voltage dependence ($V_{0.5} = -30.0$ mV, $k = -9.4$, $n = 6$). However, coexpression of $\alpha 1E$ with β_{1b} can shift the voltage dependence of inactivation 15–20 mV towards more negative potentials⁶.

One of the defining features of low-voltage-activated Ca^{2+} channels is that their unitary conductance is tiny^{1,2}. Measurement of this conductance is complicated by the low probability of channel opening at negative potentials where the driving force is stronger. Therefore, we used tail current protocols to enhance channel opening at negative potentials. Figure 4e shows representative channel openings from a single patch. Opening and closing to half-amplitude subconductance states were observed, as reported previously for cardiac T-type channels²⁴. The ensemble current (Fig. 4d) decayed with similar kinetics ($\tau = 8$ ms) as observed in the macroscopic current. The average single-channel conductance was 7.5 ± 1.5 pS ($n = 7$ patches). An average of the values reported in the literature for neuronal T-type channels was 7.7 ± 1.1 pS ($n = 10$ studies, reviewed in ref. 4).

Our results demonstrate that we have cloned the first member of the low-voltage-activated T-type Ca^{2+} channel family. Although we concentrated on the rat brain cDNA, we have also identified human and mouse homologues. The mapping of the human *CACNA1G* to human chromosome 17q22, and the mouse locus to distal mouse chromosome 11, is consistent with conserved linkage groups²⁵. A spontaneous recessive mouse neurological mutant, teetering (*tn*), maps close to the *Cacna1g* locus²⁵. Homozygous *tn/tn* mutants show brainstem and spinal cord dysgenesis, ataxia and progressive Purkinje cell loss²⁶. We show that $\alpha 1G$ mRNA has a wide distribution in brain, with a relatively high expression in the diencephalon. Functional expression of $\alpha 1G$ in oocytes allowed us to measure its biophysical properties. Based on the following criteria we unambiguously demonstrate that $\alpha 1G$ is a T-type Ca^{2+} channel: its activation at negative test potentials, slow activation kinetics at threshold potentials, fast inactivation, slow deactivation, and its small unitary conductance. □

Methods

cDNA library screening. A rat brain cDNA library (Clontech) was screened using conventional filter hybridization according to the manufacturer's protocol. Successive screening of the same cDNA library resulted in multiple overlapping clones, of which 5 were chosen to construct the full-length cDNA. The 5' terminus was synthesized by PCR (Hoffmann-La Roche Inc.) using clone D3 as template. This PCR fragment (nucleotides 5–441) was joined to a *Bam*H1 (441)–*Spe*I (1,915) fragment from clone AH3, *Spe*I–*Eag*I (3,598) from cDNA clone 5, *Eag*I–*Bgl*II (4,932) from cDNA clone 2, and *Bgl*II–*Eco*RI (7,156) from clone un7. The full-length sequence was assembled in the vector pSP72 (Promega). The sequence of $\alpha 1G$ was determined on both strands.

Northern analysis. Northern blots of 2 μ g mRNA were obtained from Clontech. The blots were hybridized at 42°C for 16–20 h in solutions containing 50% formamide. Blots were washed at temperatures up to 60°C in a final buffer of $0.1 \times$ SSC (15 mM NaCl, 1.5 mM sodium citrate), plus 0.1% SDS.

Chromosomal localization. The human chromosomal location was determined using the Genbridge 4 radiation hybrid panel and the PCR primers GACCTGAAGAAGTGCTACAG and GAGAGACTCAGCACATCTTT. PCR products were analysed by Southern hybridization with a 360 bp human genomic DNA PCR product. Fluorescent *in situ* hybridization of the *CACNA1G* cDNA clone R19524 to normal male human metaphase spreads was carried out

as described previously²⁷. The mouse chromosomal location was mapped by typing a *Sfn*N1 restriction site polymorphism in the PCR amplified product (CTCCGCCTCTGGGAGTC and ATATACATAGAGATTCTGCAC; based on AA049807).

Oocyte expression. Oocytes were prepared from *Xenopus laevis* (Nasco) using standard techniques²⁸. Each oocyte was injected with 50 nl containing 30 or 100 ng cRNA, then incubated at least four days before recording. To compare currents of similar size, we had to decrease the amount of $\alpha 1E$ injected by 1,000-fold (0.1 ng). The results were obtained from seven batches of oocytes derived from five frogs.

Electrophysiological analysis of injected oocytes. Currents were recorded using a two-microelectrode voltage-clamp amplifier as described previously²⁸. The standard bath solution contained the following: 40 mM Ba(OH)₂, 50 mM NaOH, 1 mM KOH, 0.1 mM EDTA, and 5 mM HEPES, adjusted to pH 7.4 with methanesulphonate. The osmolarity of the 2 and 10 mM Ba²⁺ solutions were balanced by increasing the NaOH concentration. Similar results were obtained in the absence of Na⁺ using BaCl₂ solutions¹⁰. Data were acquired at 4 kHz and filtered at 1 kHz, or at 10 kHz and not filtered (tail currents) using the pCLAMP system (Digidata 1200 and pCLAMP 6.0, Axon Instruments). Single channels were measured with standard depolarizing bath and pipette (115 mM BaCl₂) as described previously²⁹. Single channel data were acquired at 10 kHz, filtered on-line at 2 kHz, and filtered again off-line at 1 kHz. The data were analysed using TRANSIT³⁰. Single channel amplitudes were measured by averaging the values obtained from gaussian fits to all-points histograms of traces with openings, selected openings, or amplitude histograms of idealized openings. Boltzmann fits (curves in figures) and linear regression were calculated using Prism (GraphPad). Results are presented as mean $\pm$ s.d. unless otherwise stated.

Received 13 October; accepted 1 December 1997.

- Carbone, E. & Lux, H. D. A low voltage-activated, fully inactivating Ca channel in vertebrate sensory neurons. *Nature* **310**, 501–502 (1984).
- Nilius, B., Hess, P., Lansman, J. B. & Tsien, R. W. A novel type of cardiac calcium channel in ventricular cells. *Nature* **316**, 443–446 (1985).
- Matteson, D. R. & Armstrong, C. M. Properties of two types of calcium channels in clonal pituitary cells. *J. Gen. Physiol.* **87**, 161–182 (1986).
- Huguenard, J. R. Low threshold calcium currents in central nervous system neurons. *Annu. Rev. Physiol.* **58**, 329–348 (1996).
- Perez-Reyes, E. & Schneider, T. Molecular biology of calcium channels. *Kidney Int.* **48**, 1111–1124 (1995).
- Soong, T. W. *et al.* Structure and functional expression of a member of the low voltage-activated calcium channel family. *Science* **260**, 1133–1136 (1993).
- Williams, M. E. *et al.* Structure and functional characterization of neuronal alpha 1E calcium channel subtypes. *J. Biol. Chem.* **269**, 22347–22357 (1994).
- Wakamori, M. *et al.* Distinctive functional properties of the neuronal BII (class E) calcium channel. *Recept. Channels* **2**, 303–314 (1994).
- Schneider, T. *et al.* Molecular analysis and functional expression of the human type E $\alpha 1$ subunit. *Recept. Channels* **2**, 255–270 (1995).
- Bourinet, E. *et al.* The $\alpha 1E$ calcium channel exhibits permeation properties similar to low-voltage-activated calcium channels. *J. Neurosci.* **16**, 4983–4993 (1996).
- Wilson, R. *et al.* 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368**, 32–38 (1994).
- Jan, L. Y. & Jan, Y. N. A superfamily of ion channels. *Nature* **345**, 672 (1990).
- Stuhmer, W. *et al.* Structural parts involved in activation and inactivation of the sodium channel. *Nature* **339**, 597–603 (1989).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca²⁺ selectivity and ion permeation in L-type Ca²⁺ channels. *Nature* **366**, 158–161 (1993).
- Pragnell, M. *et al.* Calcium channel beta-subunit binds to a conserved motif in the I-II cytoplasmic linker of the alpha 1-subunit. *Nature* **368**, 67–70 (1994).
- Lambert, R. C. *et al.* T-type Ca²⁺ current properties are not modified by Ca²⁺ channel β subunit depletion in nodulus ganglion neurons. *J. Neurosci.* **17**, 6621–6628 (1997).
- de Leon, M. *et al.* Essential Ca²⁺-binding motif for Ca²⁺-sensitive inactivation of L-type Ca²⁺ channels. *Science* **270**, 1502–1506 (1995).
- Fletcher, C. F. *et al.* Absence epilepsy in tottering mutant mice is associated with calcium channel defects. *Cell* **87**, 607–617 (1996).
- Ophoff, R. A. *et al.* Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca²⁺ channel gene CACNL1A4. *Cell* **87**, 543–552 (1996).
- Tsakiridou, E., Bertolini, L., de Curtis, M., Avanzini, G. & Pape, H. C. Selective increase in T-type calcium conductance of reticular thalamic neurons in a rat model of absence epilepsy. *J. Neurosci.* **15**, 3110–3117 (1995).
- Breen, M. *et al.* Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cm resolution. *Hum. Mol. Genet.* **3**, 621–627 (1994).
- Randall, A. D. & Tsien, R. W. Contrasting biophysical and pharmacological properties of T-type and R-type calcium channels. *Neuropharmacology* **36**, 879–893 (1997).
- Shuba, Y. M., Teslenko, V. I., Savchenko, A. N. & Pogorelaya, N. H. The effect of permeant ions on single calcium channel activation in mouse neuroblastoma cells: ion-channel interactions. *J. Physiol.* **443**, 25–44 (1991).
- Droogmans, G. & Nilius, B. Kinetic properties of the cardiac T-type calcium channel in the guinea-pig. *J. Physiol.* **419**, 627–650 (1989).
- Montgomery, J. C., Silverman, K. A. & Buchberg, A. M. Chromosome 11. *Mamm. Genome* **7**, S190–208 (1997).
- Meier, M. The neuropathy of teetering, a neurological mutation in the mouse. *Arch. Neurol.* **16**, 59–66 (1967).

27. Banfi, S. *et al.* Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searches. *Nature Genet.* **13**, 167–174 (1996).
28. Bernal, J., Lee, J.-H., Cribbs, L. L. & Perez-Reyes, E. Full reversal of Pb^{2+} block of L-type Ca^{2+} channels requires treatment with heavy metal antidotes. *J. Pharmacol. Exp. Ther.* **282**, 172–180 (1997).
29. Lacerda, A. E., Perez-Reyes, E., Wei, X., Castellano, A. & Brown, A. M. T-type and N-type calcium channels of *Xenopus* oocytes: evidence for specific interactions with beta subunits. *Biophys. J.* **66**, 1833–1843 (1994).
30. VanDongen, A. M. J. A new algorithm for idealizing single ion channel data containing multiple unknown conductance levels. *Biophys. J.* **70**, 1303–1315 (1996).

Acknowledgements. We thank J. Yang for expert technical support. We also thank the UK HGMP Resource Centre for the Genbridge 4 radiation hybrid mapping panel and the EUCIB mouse panel. This work was supported by grants from the NIH, Potts Foundation and the Medical Research Council. E.P.-R. is an established investigator of the American Heart Association.

Correspondence and requests for materials should be addressed to E.P.-R. (e-mail: eperez@luc.edu). The α 1G sequences have the following Genbank accession numbers: rat, AF027984; human, AF029228 and AF029229. Human gene results are at www.hgmp.mrc.ac.uk/cgi-bin/contig/rhmapper.pl. Mouse gene results are coded as 97/MW/001 at www.hgmp.mrc.ac.uk/Mbx/MbxHomepage.html.

Disruption of IRS-2 causes type 2 diabetes in mice

Dominic J. Withers^{*,†}, Julio Sanchez Gutierrez^{*,†}, Heather Towery^{*}, Deborah J. Burks^{*}, Jian-Ming Ren[‡], Stephen Previs[‡], Yitao Zhang^{*}, Dolores Bernal^{*}, Sebastian Pons^{*}, Gerald I. Shulman[‡], Susan Bonner-Weir^{*} & Morris F. White^{*}

^{*} Howard Hughes Medical Institute, Joslin Diabetes Center, Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA

[‡] Howard Hughes Medical Institute, Department of Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA

[†] These authors contributed equally to this work.

Human type 2 diabetes is characterized by defects in both insulin action and insulin secretion. It has been difficult to identify a single molecular abnormality underlying these features. Insulin-receptor substrates (IRS proteins) may be involved in type 2 diabetes: they mediate pleiotropic signals initiated by receptors for insulin and other cytokines¹. Disruption of IRS-1 in mice retards growth, but diabetes does not develop because insulin secretion increases to compensate for the mild resistance to insulin^{2,3}. Here we show that disruption of IRS-2 impairs both peripheral insulin signalling and pancreatic β -cell function. IRS-2-deficient mice show progressive deterioration of glucose homeostasis because of insulin resistance in the liver and skeletal muscle and a lack of β -cell compensation for this insulin resistance. Our results indicate that dysfunction of IRS-2 may contribute to the pathophysiology of human type 2 diabetes.

Secretion of insulin from pancreatic β -cells tightly regulates glucose homeostasis by stimulating use of glucose by peripheral tissues and inhibiting hepatic glucose production⁴. Development of type 2 diabetes is characterized by a particular breakdown of this system: hyperinsulinaemia compensates for the resistance to insulin that is found in the early prediabetic state⁵. The subsequent development of hyperglycaemia results from the failure of β -cells to secrete enough insulin for effective compensation⁵. The cellular response to insulin is mediated by tyrosine phosphorylation of several cytosolic docking proteins (IRS proteins), which couple the insulin receptor to various effector molecules, including phosphatidylinositol-3-OH kinase (PI(3)K), Grb2/SOS, SHP2, NCK and CRK¹. Identification of IRS-2 (ref. 6) as an alternative insulin receptor substrate in IRS-1-deficient mice⁷ indicates that this may have an important role in mediating glucose homeostasis. Therefore, to characterize the role of IRS-2 in insulin-mediated regulation of glucose metabolism, we inactivated this gene by homologous recombination (Fig. 1a).

Heterozygous *IRS-2*^{+/-} offspring survived to adulthood and were interbred to homozygosity. Mice lacking IRS-2 were identified by Southern blot analysis, and the absence of IRS-2 was confirmed by

western blotting of liver and skeletal muscle tissue extracts (Fig. 1b, c). Analysis of IRS expression in these tissues showed comparable levels of IRS-1 in *IRS-2*^{-/-} and wild-type animals and comparable levels of IRS-2 in *IRS-1*^{-/-} and wild-type mice (Fig. 1d). *IRS-2*^{-/-} neonates were 10% smaller than *IRS-2*^{+/-} or wild-type littermates, and this small difference in weight persisted during weaning and into adult life (Fig. 1e).

Blood sugar levels were analysed at various ages to establish the role of IRS-2 in glucose homeostasis. Three days after birth, levels of randomly fed sugars were elevated in *IRS-2*^{-/-} mice (166 ± 13 mg dl⁻¹) compared with in wild-type mice (116 ± 10 mg dl⁻¹). Monitoring of blood glucose between 3 and 6 weeks of age showed the development of fasting hyperglycaemia in *IRS-2*^{-/-} mice, and at 6–8 weeks of age these mice exhibited marked glucose intolerance during an intraperitoneal glucose-tolerance test (Fig. 2a, b). At 10 weeks, *IRS-2*^{-/-} mice were overtly diabetic with fasting glucose levels of 323 ± 35 mg dl⁻¹, and if left untreated the levels of fasting sugars rose progressively to >400 mg dl⁻¹ at 12–16 weeks of age. Male *IRS-2*^{-/-} mice showed polydipsia and polyuria

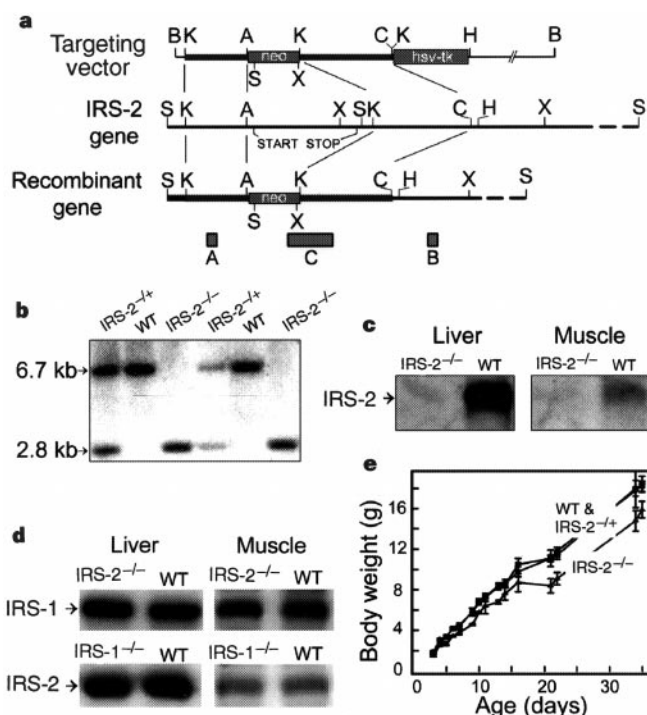


Figure 1 Gene targeting of the *IRS-2* locus. **a**, Restriction map of the targeting vector (top), the *IRS-2* gene (middle) and the homologous recombinant gene (bottom). Targeted deletion of the entire *IRS-2* gene and its replacement with the PGK/neo cassette was done using the targeting vector. Restriction enzymes are as follows: A, *Apa*1; B, *Bcl*1; C, *Cla*1; H, *Hind*III; K, *Kpn*1; S, *Spe*1; X, *Xma*1. Probes A–C were used to screen for homologous recombination by Southern blotting. **b**, Southern-blot analysis from homozygous mutant mice (*IRS-2*^{-/-}), heterozygotes (*IRS-2*^{+/-}) and wild-type animals (WT). Tail DNA was digested with *Spe*1 and analysed by Southern blotting with a 0.3-kilobase (kb) external genomic fragment (probe A). A 6.7-kb fragment is indicative of the wild-type allele and a 2.8-kb fragment is indicative of the recombinant allele. **c**, Immunoprecipitation and western-blot analysis of liver and muscle, showing lack of IRS-2 protein in *IRS-2*^{-/-} mice. Each lane represents tissue from one animal and represents three independent experiments. **d**, Immunoprecipitation and western-blot analysis of the liver and muscle expression of IRS-1 in wild-type and *IRS-2*^{-/-} mice, and of IRS-2 in wild-type and *IRS-1*^{-/-} mice. Each lane represents tissue from one animal and represents three independent experiments. **e**, Mean weight $\pm$ s.e.m. of WT, *IRS-2*^{+/-} and *IRS-2*^{-/-} animals on a C57B16 $\times$ 129/Sv background at the indicated ages (days). Data are from six litters, with a total of at least 15 animals per genotype.

without ketosis, and died from dehydration and hyperosmolar coma; female mice followed a similar disease progression but rarely died. Thus, deletion of *IRS-2* progressively impaired glucose tolerance with greater than 95% penetrance, whereas heterozygous and wild-type animals were unaffected.

The progressive development of diabetes in *IRS-2*-deficient mice suggested that deletion of this gene caused a combination of peripheral insulin resistance and inadequate compensatory insulin secretion because of relative β -cell failure. To determine whether *IRS-2*^{-/-} mice were insulin-resistant, we measured fasting serum insulin levels and determined insulin sensitivity *in vivo*. Insulin levels in wild-type and *IRS-2*^{-/-} neonates were not significantly different, suggesting that *IRS-2*^{-/-} mice were slightly insulin-resistant in light of their increased random glucose levels, mentioned above. However, after 6 weeks *IRS-2*^{-/-} mice exhibited threefold higher fasting insulin levels than wild-type animals, and insulin-tolerance tests showed a significantly reduced hypoglycaemic response to exogenous insulin (Fig. 2c, d).

To define the nature of the insulin resistance, we determined basal and insulin-stimulated whole-body glucose disposal in conscious mice using the euglycaemic hyperinsulinaemic clamp and tracer

techniques⁸. Six-to-eight-week-old wild-type and *IRS-2*^{-/-} mice in the fasted state showed identical basal rates of glucose disposal and levels of production of hepatic glucose, despite the presence of hyperinsulinaemia in the *IRS-2*^{-/-} animals. However, low-dose insulin infused at a rate of 2.5 mU kg⁻¹ min⁻¹ did not increase whole-body glucose disposal or suppress hepatic glucose release in *IRS-2*^{-/-} mice, although this dose of insulin markedly enhanced these effects in wild-type and *IRS-2*^{+/-} mice (Fig. 2e, f). A higher insulin dose (20 mU kg⁻¹ min⁻¹) increased glucose disposal and suppressed hepatic glucose production in *IRS-2*^{-/-} mice, confirming that there was profound insulin resistance in both skeletal muscle and liver.

IRS proteins are tyrosine-phosphorylated by the insulin receptor

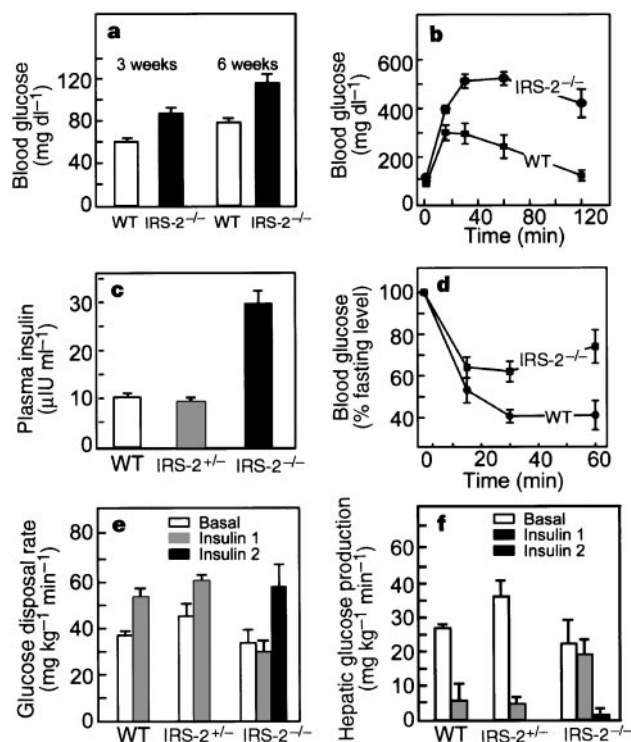


Figure 2 Fasting blood glucose and glucose-tolerance test, fasting insulin levels and insulin-tolerance test, and *in vivo* glucose disposal and hepatic glucose production. **a**, After a 15 h overnight fast, blood glucose levels were determined using a Glucometer Elite glucometer (Bayer). Results are mean values $\pm$ s.e.m. for at least eight animals per genotype, with ages as indicated. WT, wild type. **b**, Glucose-tolerance tests after intraperitoneal loading with 2 g D-glucose per kg were performed on 6-week-old animals of the indicated genotype. Results are mean values $\pm$ s.e.m. for at least eight animals per genotype. **c**, Serum insulin levels were measured by radioimmunoassay on 4–6-week-old anaesthetized animals after a 15 h overnight fast. Data are the mean values $\pm$ s.e.m. for at least 12 animals per genotype. **d**, Insulin-tolerance tests were performed on fed 4–6-week-old animals. Results are expressed as percentage of initial blood glucose concentration and are the mean values $\pm$ s.e.m. for at least eight animals per genotype. **e**, Glucose-disposal rate and **f** hepatic glucose production rate were determined on fasted, conscious 8-week-old mice using the euglycaemic hyperinsulinaemic clamp. Basal rates and those stimulated by infusion of insulin at a rate of 2.5 mU kg⁻¹ min⁻¹ (insulin 1) and 20 mU kg⁻¹ min⁻¹ (insulin 2) were determined. Results are the mean values $\pm$ s.e.m. for three animals per genotype.

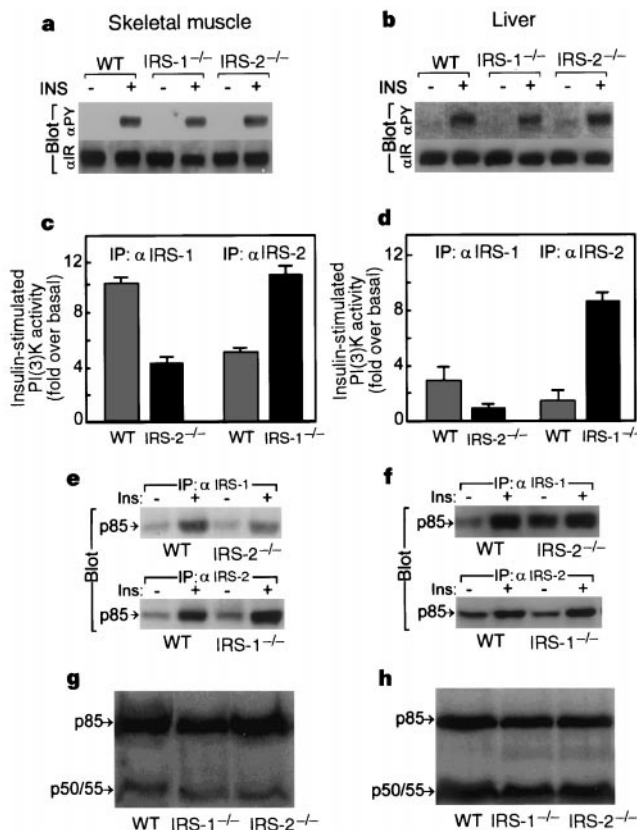


Figure 3 Expression and insulin-stimulated tyrosine phosphorylation of the insulin receptor (IR), insulin-stimulated activation of PI(3)K and IRS association with p85, and expression of PI(3)K adaptor subunits in the liver and muscle of wild-type (WT), *IRS-1*^{-/-} and *IRS-2*^{-/-} mice. Supernatants of muscle (**a**) or liver (**b**) homogenates containing equal amounts of protein from untreated and insulin (INS)-treated 4–6-week-old mice were immunoprecipitated with anti-IR β antibody and blotted for either antiphosphotyrosine (αPY) or IR β (αIR). Data are representative of data obtained from three animals per genotype. Supernatants of muscle (**c**) or liver (**d**) homogenates containing equal amounts of protein from untreated and insulin-treated 4–6-week-old mice were immunoprecipitated (IP) in duplicate with the indicated antibody (αIRS-1 or αIRS-2); immunoprecipitates were assayed *in vitro* for PI(3)K activity. Data are the mean values $\pm$ s.e.m. of two independent experiments and represent data from a total of eight wild-type, nine *IRS-2*^{-/-} and four *IRS-1*^{-/-} animals. Results are expressed as fold stimulation of activity above that of non-insulin-treated controls (fold stimulation over basal). Muscle (**e**) and liver (**f**) homogenates treated as above were immunoprecipitated with the indicated antibody (αIRS-1 or αIRS-2) and subjected to western analysis to study association with p85 α/β . Data are representative of data obtained from three animals per genotype. Ins, insulin. Muscle (**g**) and liver (**h**) lysates from animals of the indicated genotypes were subjected to SDS-PAGE and western blotting with an anti-p85 SH2 domain antiserum, which recognize p85 α/β and the p50/55 splice variants. Data are representative of those obtained from three animals per genotype.

during insulin stimulation and bind to the Src homology 2(SH2) domains in various effector proteins, including PI(3)K (ref. 9). We examined the proximal steps in this signalling cascade in *IRS-2*^{-/-}, *IRS-1*^{-/-} and wild-type mice to determine the potential contribution of any defects in these signalling elements to the development of peripheral insulin resistance. Equivalent insulin-receptor expression and tyrosine phosphorylation of the insulin-receptor β -subunit were seen in both the muscle and liver of *IRS-2*^{-/-} and wild-type mice (Fig. 3a, b). PI(3)K may be involved in mediating several insulin-regulated metabolic pathways, including glucose uptake¹⁰, antilipolysis¹¹, glycogen synthesis¹² and the suppression of hepatic gluconeogenesis through the regulation of phosphoenolpyruvate carboxykinase (PEPCK) expression¹³. Abnormalities in the activation of PI(3)K could explain defective glucose homeostasis in the *IRS-2*^{-/-} mice. Therefore, the increase in PI(3)K activity after stimulation with insulin of IRS-1 immunoprecipitates from the muscle and liver of *IRS-2*^{-/-} mice was compared with the increase in PI(3)K activity in wild-type animals and contrasted with the PI(3)K activity detected in IRS-2 immunoprecipitates from these tissues from *IRS-1*^{-/-} mice and wild-type mice. The increase in PI(3)K activity associated with IRS-1 upon stimulation with insulin was reduced by >50% in both muscle and liver of *IRS-2*^{-/-} mice as compared with wild-type animals (Fig. 3c, d). The reduction in fold stimulation was due in part to an increase in basal PI(3)K activity in IRS-1 immunoprecipitates from liver and muscle of *IRS-2*^{-/-} mice. These findings suggest a potential defect in the ability of cells to appropriately regulate both basal and insulin-stimulated PI(3)K activity in the absence of IRS-2.

In contrast, the insulin-stimulated PI(3)K activity in IRS-2

immunoprecipitates from the liver and muscle of *IRS-1*^{-/-} animals was markedly enhanced when compared with the IRS-2-associated PI(3)K activity in the tissues of wild-type animals, as previously reported⁷. This pattern of PI(3)K activation in muscle and liver paralleled the insulin-stimulated association of the p85 subunit of PI(3)K with IRS-1 in wild-type and *IRS-2*^{-/-} animals and its association with IRS-2 in wild-type and *IRS-1*^{-/-} mice (Fig. 3e, f). However, no differences in expression of p85 α/β , p55^{PIK} and the splice variants p50/55 were detected in the liver and muscle of animals of the three genotypes (Fig. 3g, h). Therefore, the functional defects observed in insulin-stimulated PI(3)K activation in *IRS-2*^{-/-} mice may underlie the abnormalities in glucose metabolism in these animals.

Although insulin resistance is important in the early stages of type 2 diabetes in humans, the failure in adequate β -cell compensation leads to the progression to the diabetic state⁵. Compensation for insulin resistance can be achieved either by greater insulin secretion per β -cell or by an increase in β -cell mass through neogenesis or replication of the existing β -cells¹⁴. Morphometric analysis of pancreases from mice at 4 weeks of age, a time when there is normally a significant increase in β -cell mass¹⁵, showed that *IRS-2*^{-/-} mice had significantly reduced β -cell mass (0.278 ± 0.04 mg) compared with wild-type mice (0.677 ± 0.09 mg), but no significant difference in non- β endocrine-cell mass (Fig. 4a, b, d and data not shown). In contrast, the β -cell mass of *IRS-1*^{-/-} mice, which have insulin resistance without diabetes, was almost double that of wild-type mice (1.280 ± 0.07 mg) (Fig. 4c, d). Subsequent examination of neonatal *IRS-2*^{-/-} pancreas showed relative β -cell deficiency, suggesting that these changes are independent of long-term metabolic effects (data not shown). Functional assessment of insulin release *in vivo* during a glucose-tolerance test showed that 4-week-old wild-type mice (with a fasting glucose level of 79 ± 5 mg dl⁻¹) exhibited a twofold increase in circulating insulin levels (from 11.75 ± 1.4 international microunits (μ IU) ml⁻¹ to 24.4 ± 2.7 μ IU ml⁻¹; $n = 4$) 60 minutes after glucose loading. Similarly, *IRS-2*^{-/-} mice (with fasting glucose levels of 106 ± 5 mg dl⁻¹), despite fasting hyperinsulinemia, responded with a 1.9-fold increase in insulin levels 60 minutes after glucose loading (fasting, 23.5 ± 4.7 μ IU ml⁻¹; 60 min, 42.1 ± 5.8 μ IU ml⁻¹; $n = 4$). These results indicate that, in the early stages of the development of the diabetic phenotype, glucose-stimulated insulin release may be nearly normal. However, as the hyperglycaemia progresses we see attenuation of glucose-stimulated insulin release, which may be attributed to glucose toxicity towards β -cell function (data not shown).

To obtain further insights into the role of IRS-2 in β -cell function, we studied the expression of this protein in the islets of wild-type animals. Immunofluorescence staining showed that IRS-2 expression co-localized with insulin in the islets, indicating that it may be present in β -cells not non- β cells (Fig. 4a–g). There was also significant IRS-2 staining in the ductal epithelium, which is the site of neogenesis of new islets from ductal precursor cells¹⁴. IRS-2 staining was absent from the islets and ducts of *IRS-2*^{-/-} mice (Fig. 4h, i). Thus, IRS-2-dependent signalling pathways may be important for the cells that are involved in the proliferative and neogenic responses of the islets.

Our results indicate that deletion of IRS-2 causes the progressive development of a type 2 diabetic phenotype in mice. *IRS-2*^{-/-} mice exhibit mild peripheral insulin resistance and β -cell deficiency at birth but have adequate compensatory insulin secretion for several weeks. However, subsequent relative β -cell failure in the face of continued peripheral resistance causes overt fasting hyperglycaemia without ketoacidosis, the common characteristic of human type 2 diabetes (ref. 5). The mechanisms of peripheral insulin resistance are unknown at present, and the exact contribution of the observed insulin resistance of muscle and liver to the progression of the disease requires further study.

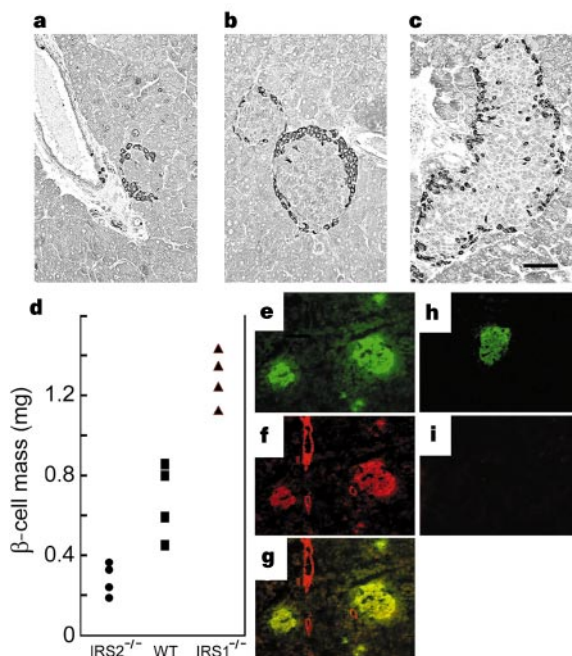


Figure 4 Islet morphology and analysis of β -cell mass in *IRS-2*^{-/-}, *IRS-1*^{-/-} and wild-type mice, and expression of IRS-2 in islets of wild-type and *IRS-2*^{-/-} mice. **a–c**, Immunostaining for non- β -cell hormones of pancreas sections from 4-week-old male animals. Representative islets from **(a)** *IRS-2*^{-/-}, **(b)** wild-type and **(c)** *IRS-1*^{-/-} animals are shown and demonstrate the relative differences in islet-cell mass between the animals of each genotype (bar represents 100 μ m). **d**, Quantification of β -cell mass by point-counting morphometric analysis on sections from the same mice analysed blind for the genotype. Mean total pancreatic weights (g) $\pm$ s.e.m. for each genotype were: wild-type, 178 ± 19 ; *IRS-2*^{-/-}, 123 ± 12 ; *IRS-1*^{-/-}, 113 ± 9 ($n = 4$). **e–g**, Immunostaining in wild-type animals for **(e)** insulin, **(f)** IRS-2, and colocalization of both proteins **(g)**. **h, i**, Immunostaining in *IRS-2*^{-/-} animals for insulin **(h)** and IRS-2 **(i)**. Representative fields showing both islets and pancreatic ducts are shown.

Our results provide insight into the potential differences in the physiological roles of IRS-1 and IRS-2. *IRS-2^{-/-}* mice show marked abnormalities in glucose homeostasis but minimal growth defects, whereas the opposite is the case for *IRS-1^{-/-}* mice. Thus, IRS-1 and IRS-2 cannot be functionally interchanged to produce either IGF-1-stimulated mitogenesis, as has been suggested by previous *in vitro* data¹⁶, or, as we demonstrate, to produce insulin-regulated metabolism. The signalling specificity through IRS-1 and IRS-2 may be accomplished by specific expression patterns and distinct phosphorylation patterns during interaction with various activated receptors¹⁷. Our observations of PI(3)K activity in muscle and liver show functional differences in the ability of IRS proteins to regulate PI(3)K and implicate IRS-2 as the more critical *in vivo* regulator of this signalling pathway, which mediates many of the metabolic effects of insulin.

The combination of reduced β -cell mass and a failure of islet hyperplasia in the face of insulin resistance and hyperglycaemia distinguishes the *IRS-2^{-/-}* mouse from other monogenic and polygenic models of type 2 diabetes. For example, *IRS-1^{-/-}* mice, or mice with compound heterozygous disruptions of IRS-1 and the insulin receptor¹⁸, develop marked β -cell hyperplasia in response to insulin resistance; this is not seen in human lean type 2 diabetics¹⁹. Likewise, *IRS-2^{-/-}* mice cannot compensate for insulin resistance in this manner. Our results indicate a unique role for IRS-2 in the regulation of β -cell neogenesis, proliferation and survival. The progressive nature of the diabetes in these mice is due to both insulin resistance and reduced β -cell mass, which prevents adequate compensation. As this combination of features is the hallmark of human type 2 diabetes, functional abnormalities in IRS-2 could be involved in the pathogenesis of this human disease. □

Methods

Preparation of the construct for homologous recombination and generation of IRS-2-deficient mice. We cloned the IRS-2 gene from a 129 mouse genomic library¹⁷. To construct the targeting vector, two fragments of the genomic DNA flanking the coding region were subcloned at convenient restriction sites into the pPNT vector. We transfected linearized pPNT/IRS-2 into the R1 line of embryonic stem (ES) cells derived from 129 mouse blastocysts. Selection was performed with G418 and ganciclovir, and resistant clones were screened for homologous recombination by Southern blotting using 5' and 3' external probes (probes A and B) and by an internal probe (C) derived from the neomycin-resistance gene (*neo*) cassette (Fig. 1a). One cell clone fulfilled the requirements for homologous recombination. Blastocysts from C57B1/6 mice were injected with 10–16 targeted ES cells and implanted into pseudopregnant CD-1 foster mothers as described²⁰. Several chimaeric male pups were obtained and mated with C57B1/6 females. Germline transmission was confirmed by Southern blotting and heterozygote offspring were intercrossed to obtain animals homozygous for the deletion of IRS-2. These animals were used in subsequent studies with the animals maintained with a C57B1/129sv hybrid background. The IRS-1 targeting vector was constructed in a similar manner and *IRS-1^{-/-}* animals were maintained on the same genetic background as the IRS-2 animals to facilitate comparison of the *IRS-1^{-/-}* and *IRS-2^{-/-}* phenotypes.

Metabolic studies. Animals were maintained on a normal light/dark cycle and handled in accordance with Joslin Diabetes Center Animal Care and Use Committee protocols. Glucose levels were determined from blood taken from mouse tails using a Glucometer Elite glucometer (Bayer). Blood for plasma insulin levels was taken either by retroorbital bleeds from anaesthetized mice or by tail bleeds. Immunoreactive insulin levels were measured by radioimmunoassay using rat insulin (Linco) as a standard. Glucose-tolerance tests were performed on animals after a 15-h overnight fast. Animals were injected with 2 g kg⁻¹ of D-glucose intraperitoneally and blood glucose values were determined at the times indicated. Insulin tolerance was tested with fed animals between 14:00 and 16:00. Animals were injected with 0.75 units per kg body weight with human crystalline insulin (Lilly) intraperitoneally. Blood was taken immediately before injection and at the times indicated. Results were expressed as percentages of initial blood glucose concentration.

Euglycaemic hyperinsulinaemic clamps were performed on fasted conscious mice using a two-step clamp technique⁸. [³H]glucose was infused throughout the clamp study to determine glucose-turnover rate. After a priming dose, [³H]glucose was continuously infused at a rate of <0.08 μ Ci min⁻¹ for 5 h. For basal glucose turnover rate measurements, blood samples were collected at 70 and 80 min after the initiation of [³H]glucose infusion. Insulin (Eli Lilly) was infused at a rate of 2.5 mU kg⁻¹ for 90 min while 20% dextrose was infused by variable infusion pump. Blood samples were collected from tail-tip bleeds for glucose estimation every 10 min. Plasma glucose was clamped at 100 mg dl⁻¹. While glucose levels remained steady, two blood samples were taken for [³H]glucose-specific-activity determination. The insulin infusion rate was then increased to 20 mU kg⁻¹ min⁻¹ for 90 min. Analysis of [³H]glucose measurements and calculation of glucose-disposal rates and hepatic glucose production rates were performed as described⁸.

Immunoprecipitations, western blotting and PI(3)K assays. Liver and muscle tissue lysates were removed and homogenized at 4 °C as described⁷. The homogenates were solubilized for 1 h at 4 °C and clarified by centrifugation at 15,000 r.p.m. for 30 min. Supernatants containing equal amounts of protein were immunoprecipitated for 2 h with an anti-IRS-2 antibody raised against a glutathione-S-transferase (GST)-fusion protein containing residues 619–746 of murine IRS-2, anti-IRS-1 antibody raised against a GST-fusion protein containing residues 735–900 of murine IRS-1, or an antibody against insulin receptor subunit β . Immune complexes were collected with 100 μ l of a 50% slurry of protein-A-sepharose resolved on 7.5% SDS-PAGE and transferred to nitrocellulose. The blots were probed with polyclonal antibodies against IRS-1 and IRS-2 and anti-p85 SH2 domain, which recognizes p85 α/β , p55^{PIK} and the p50/55 splice variants²¹. Subsequent detection was by either [¹²⁵I]protein A or enhanced chemiluminescence. For assays of p85 association with IRS proteins and PI(3)K enzymatic activity, 5 units of human insulin were injected as a bolus into the inferior vena cava of anaesthetized mice and the liver, gastrocnemius and quadriceps muscles were removed at 1, 2.5 and 3 min after insulin injection. They were homogenized and solubilized, and supernatants containing equal amounts of protein were immunoprecipitated for 2 h with anti-IRS-2 or with anti-IRS-1C antibodies. Immune complexes were collected and washed extensively and the PI(3)K reaction was performed as described⁷. [³²P]incorporation was quantified using a Phosphorimager (Molecular Dynamics).

Immunocytochemistry and immunofluorescence. Animals were killed by administering an overdose of sodium amytal. The pancreases were removed, cleared of fat and lymph nodes, weighed, and, for immunocytochemistry, were fixed in Bouin's solution and embedded in paraffin. Sections (of 5 μ m) were immunostained for the endocrine non- β -cells, using a cocktail of antibodies (rabbit antibodies against bovine glucagon, against synthetic somatostatin and rabbit against bovine pancreatic polypeptide¹⁵). β -cell mass was determined by point-counting morphometry as described²² at a final magnification of $\times$ 420, with at least 175 fields being quantified per animal. For immunofluorescence, pancreases were snap-frozen, and 5- μ m cryosections were cut and fixed in paraformaldehyde. Sections were permeabilized with 0.2% Triton X-100 and stained with anti-IRS-2 antisera and guinea pig antibodies against porcine insulin, and detection was performed with rhodamine- and fluorescein-conjugated secondary antibodies.

Received 24 November; accepted 29 December 1997.

- Myers, M. G. Jr & White, M. F. Insulin signal transduction and the IRS proteins. *Annu. Rev. Pharmacol. Toxicol.* **36**, 615–658 (1996).
- Araki, E., Lipes, M. A., Patti, M. E., Bruning, J. C., Haag, B. L. III, Johnson, R. S. & Kahn, C. R. Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **372**, 186–190 (1994).
- Tamamoto, H. et al. Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1. *Nature* **372**, 182–186 (1994).
- Warram, J. H., Rich, S. S. & Krolewski, A. S. *Joslin's Diabetes Mellitus* (eds Kahn, C. R. & Weir, G. C.) 201–215 (Philadelphia, Lea & Febiger, 1994).
- DeFronzo, R. A., Bonadonna, R. C. & Ferrannini, E. Pathogenesis of NIDDM: a balanced overview. *Diabetes Care* **15**, 318–368 (1992).
- Sun, X. J. et al. Role of IRS-2 in insulin an cytokine signalling. *Nature* **377**, 173–177 (1995).
- Patti, M. E. et al. 4PS/IRS-2 is the alternative substrate of the insulin receptor in IRS-1 deficient mice. *J. Biol. Chem.* **270**, 24670–24673 (1995).
- Ren, J. M. et al. Overexpression of Glut4 protein in muscle increases basal and insulin-stimulated whole body glucose disposal in conscious mice. *J. Clin. Invest.* **95**, 429–432 (1995).
- Backer, J. M. et al. Association of IRS-1 with the insulin receptor and the phosphatidylinositol 3'-kinase. Formation of binary and ternary signaling complexes in intact cells. *J. Biol. Chem.* **268**, 8204–8212 (1993).
- Cheatham, B. et al. Phosphatidylinositol 3-kinase activation is required for insulin stimulation of pp70 S6 kinase DNA synthesis and glucose transporter translocation. *Mol. Cell Biol.* **14**, 4902–4911 (1994).

11. Okada, T., Kawano, Y., Sakakibara, T., Hazeki, O. & Ui, M. Essential role of phosphatidylinositol 3-kinase in insulin-induced glucose transport and antilipolysis in rat adipocytes: studies with a selective inhibitor wortmannin. *J. Biol. Chem.* **269**, 3568–3573 (1994).
12. Shepherd, P. R., Nave, B. T. & Siddle, K. Insulin stimulation of glycogen synthesis and glycogen synthase activity is blocked by wortmannin and rapamycin in 3T3-L1 adipocytes: evidence for the involvement of phosphoinositide 3-kinase and p70 ribosomal protein-S6 kinase. *Biochem. J.* **305**, 25–28 (1995).
13. Gabbay, R. A. *et al.* Insulin regulation of phosphoenolpyruvate carboxykinase gene expression does not require activation of the Ras/mitogen-activated protein kinase signaling pathway. *J. Biol. Chem.* **271**, 1890–1897 (1996).
14. Bonner-Weir, S., Scaglia, L., Montana, E., Juang, J. H. & Weir, G. C. in *Diabetes 1994* (eds Baba, S. & Kareko, T.) 179–228 (Excerpta Medica Int. Congress, 1995).
15. Scaglia, L., Cahill, C. J., Finegood, D. T. & Bonner-Weir, S. Apoptosis participates in the remodeling of the endocrine pancreas in the neonatal rat. *Endocrinology* **138**, 1736–1741 (1997).
16. Bruning, J. C., Winnay, J., Cheatham, B. & Kahn, C. R. Differential signaling by insulin receptor substrate 1 (IRS-1) and IRS-2 in IRS-1 deficient cells. *Mol. Cell Biol.* **17**, 1513–1521 (1997).
17. Sun, X. J. *et al.* The IRS-2 gene on murine chromosome 8 encodes a unique signaling adapter for insulin and cytokine action. *Mol. Endocrinol.* **11**, 251–262 (1997).
18. Bruning, J. C. *et al.* Development of a novel polygenic model of NIDDM in mice heterozygous for *IR* and *IRS-1* null alleles. *Cell* **88**, 561–572 (1997).
19. Kloppel, G., Lohr, M., Habich, K., Oberholzer, M., Heitz, P. U. Islet pathology and the pathogenesis of type 1 and type 2 diabetes mellitus revisited. *Surv. Synth. Pathol. Res.* **4**, 110–125 (1985).
20. Papaioannou, V., Johnson, R. in *Gene Targeting: A Practical Approach* (ed. Joyner, A.) Ch. 4, pp. 107–146 (Oxford Univ. Press, Oxford, 1993).
21. Pons, S. *et al.* The structure and function of p55PIK reveals a new regulatory subunit for the phosphatidylinositol-3 kinase. *Mol. Cell Biol.* **15**, 4453–4465 (1994).
22. Montana, E., Bonner-Weir, S. & Weir, G. C. Transplanted beta cell response to increased metabolic demand. Changes in beta cell replication and mass. *J. Clin. Invest.* **93**, 1577–1582 (1994).

Acknowledgements. We thank A. Nagy for R1 cells; M. Ginsberg, M. Petrucci and M. Taneja for technical support; B. Cheatham for the anti-IRB antibody. Blastocyst injections were performed in the Core Laboratory of the Diabetes and Endocrinology Research Center, Vanderbilt University. This work was supported by grants from the NIH to G.I.S., S.B.W. and M.E.W. D.J.W. is supported by an MRC (UK) Clinician Scientist Fellowship. D.J.B. was supported by a grant from the JDFI. J.S.G. and D.B. were supported by grants from the Spanish Government.

Correspondence and requests for materials should be addressed to M.E.W. (e-mail: Whittemor@joslab.harvard.edu).

Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor

Anne E. Corcoran, Andrew Riddell, Danielle Krooshoop & Ashok R. Venkitaraman

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

To generate the full diversity of antibody heavy-chain genes, hundreds of dispersed germline *V* segments must undergo recombination following *D–J* segment joining. Here we report that this process is regulated by the α -chain of the receptor for interleukin-7, a cytokine that stimulates B-cell lymphopoiesis¹. *D–J* joining occurs normally in immature B lymphocytes from mice lacking the α -chain of the interleukin-7 receptor (IL-7R α). But recombination of *V* segments is progressively impaired as their distance increases upstream of *D/J*, causing infrequent rearrangement of most *V* segments, which markedly reduces diversity. This is not simply due to defective cell proliferation or impaired recombinase expression. Rather, germline transcripts from distal, unrearranged *V* segments, a marker of chromatin changes that precede recombination, are specifically silenced. So too is expression of Pax-5, which binds to heavy-chain locus control elements and normally stimulates recombination, suggesting a mechanism for these effects. Thus ligands of the interleukin-7 receptor deliver an extrinsic signal that targets *V* segment recombination in the heavy-chain locus by altering the accessibility of DNA substrates to the recombinase. This mechanism augments the recombinational diversity of the primary antibody repertoire.

B lymphopoiesis is impeded at an early stage in the bone marrow of mice lacking IL-7R α ¹. The number of progenitor (pro)-B cells undergoing immunoglobulin heavy-chain gene (*IgH*) rearrangement is normal (Table 1). But there is a severe (~10-fold) reduction

in precursor (pre)-B cells with complete *IgH* rearrangements, and in surface (s) IgM⁺ B lymphocytes when compared to age-matched heterozygous controls (Table 1). This developmental block ensues not only from defective proliferation^{1,2}, but also from an impediment to *IgH* rearrangements whose basis is undetermined².

To examine *IgH* rearrangements, we purified fractions of immature B lymphocytes from IL-7R α ^{−/−} animals and $\pm$ controls. As only a few cells could be isolated by fluorescent cell sorting (Table 1), genomic DNA was analysed³ by polymerase chain reaction (PCR) to provide a semiquantitative measure of the relative frequency of recombination events. PCR primers are separated in the germ line (Fig. 1a) but are brought into sufficient proximity for PCR amplification only after a recombination event (Fig. 1b).

D–JH recombination proceeds normally without IL-7R α (Fig. 1c). There is no difference between IL-7R α ^{−/−} pro-B cells and $\pm$ controls in the frequency of (*D*)*JH* joins involving 13 of 15 known *DH* segments⁴.

In contrast, *VH–(D)JH* joining is impaired. *VH* segments in *IgH* are grouped into homologous families arranged in clusters (Fig. 1a) across approximately a megabase of DNA upstream of *D* segments and the *J* cluster. Most *VH* segments are located at the extreme 5' end of this region^{5,6} and belong to the large *VHJ558* family (with an estimated 60–1,000 members). In contrast, the small *VH7183* family containing only ~25 members is adjacent to *DH* segments at its 3' end. Rearrangements involving these *VH* segment families were distinguished by a PCR assay whose specificity was verified (Fig. 2a) using hybridomas harbouring known rearrangements. *VH* to (*D*)*JH* joins involving *VHJ558* segments are normally the most frequent rearrangements detected in the sIgM⁺ B lymphocyte fraction^{7,8}. Surprisingly, *VHJ558* rearrangements were barely detectable in IL-7R α ^{−/−}, sIgM⁺ cells compared to $\pm$ controls, despite

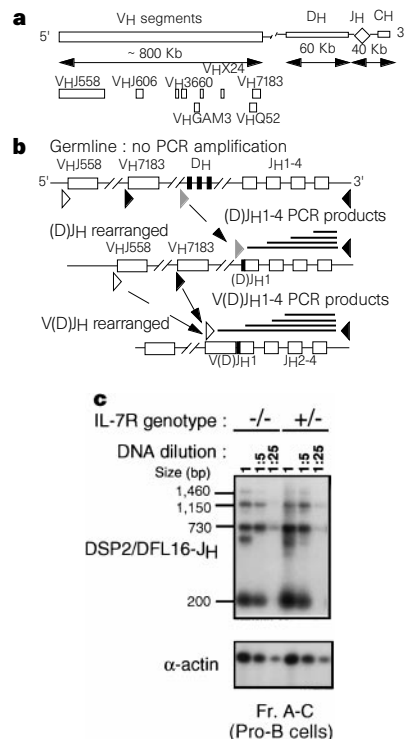


Figure 1 Normal (*D*)*JH* recombination in IL-7R α ^{−/−} mice. **a**, The murine *IgH* locus⁶, showing approximate map positions of the different *VH* families. **b**, PCR assays used to quantify (*D*)*JH* and *V* to (*D*)*JH* rearrangements. Arrows indicate specific oligonucleotide primers. **c**, There is no difference between IL-7R α ^{−/−} and IL-7R α ^{+/−} pro-B cells in the frequency of (*D*)*JH* joining. DSP2/FL16 primers used detect joins involving 13 of 15 known murine *DH* segments. Expected product sizes are indicated.

overexposure of autoradiographs (Fig. 2b), although by definition each sIgM⁺ cell must contain at least one productive IgH rearrangement. That there was no decrease in recombination of V_H7183 segments shows this defect to be specific.

This reduction cannot simply be explained by a selective failure in proliferative expansion of cells bearing productive V_H558 rearran-

gements because a similar, profound decrease in V_H558-(D)_H joining also occurs in IL-7Rα^{-/-} pro-B cells (Fig. 2c). Pro-B cells are unlikely to undergo selection because they predominantly contain non-productive V(D)_H rearrangements³, which do not encode functional heavy chains. Moreover, the number of cells within the pro-B cell fractions B and C is comparable in IL-7Rα^{-/-} and control

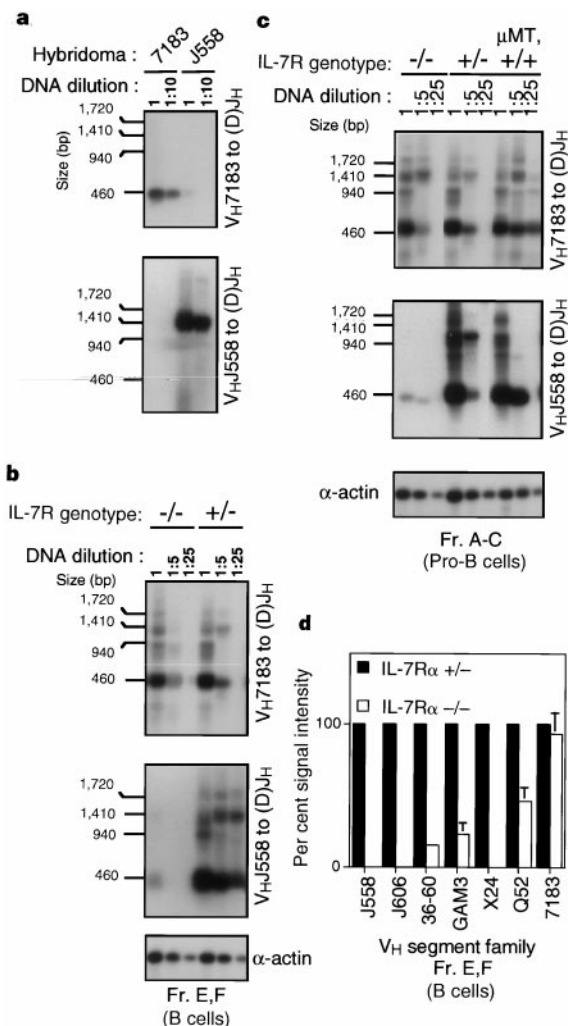


Figure 2 Impaired targeting of recombination to 5' V_H segments. **a**, Specificity of PCR assays in hybridomas (P3, for V_H7183 (ref. 30); JW1, for V_H558 (gift from M. Neuberger)). Primers used are expected on the basis of sequence homology to amplify >80% of the V_H genes listed in ref. 30). **b**, Comparison of V to (D)_H rearrangements of D_H-distal V_H558 segments with D_H-proximal V_H7183 segments in SIgM⁺ B cells. The V_H558 assay is deliberately overexposed to show the marked reduction in IL-7Rα^{-/-} cells. **c**, V_H558 and V_H7183 segment recombination in IL-7Rα^{-/-}, +/- and μMT pro-B cells. Parallel reactions with primers for a 737-bp product of the α-actin gene²⁶ serve as a quantitative control. Similar results were obtained in at least 3 independent experiments. **d**, Efficiency of V_H segment recombination in IL-7Rα^{-/-} purified sIgM⁺ B cells decreases with increasing distance from D_H/J_H. Recombination involving the indicated V_H segment families was quantified by densitometry of autoradiographs. Relative intensity of the hybridization signal corresponding to V(D)_H4 rearrangements in each family (normalized to α-actin) are plotted on the y-axis as a percentage of the signal in +/- cells. The V_H family-specific oligonucleotide primers used are described under Methods.

Table 1 Cell numbers in fractions of bone marrow from IL-7Rα^{-/-} and IL-7Rα^{+/+} mice

IL-7R genotype	N	Total cells (×10 ⁶)	B220+ cells (×10 ⁵)	Pro-B cells (fr. A-C) (×10 ⁵)	Pre-B cells (fr. D) (×10 ⁵)	B cells (fr. E,F) (×10 ⁵)
+/-	3	200 ± 15	42 ± 9.5	4 ± 1	20 ± 4.5	18 ± 6
-/-	3	120 ± 9	7 ± 2.5	3.5 ± 1.5	2.5 ± 2	1 ± 1

IL-7R genotype	N	Pro-B cells (fr. A) (×10 ⁵)	Pro-B cells (fr. B) (×10 ⁵)	Pro-B cells (fr. C) (×10 ⁵)
+/-	2	0.48 ± 0.3	2.94 ± 0.8	0.72 ± 0.2
-/-	2	0.92 ± 0.5	2.1 ± 1	0.67 ± 0.4

Fractions²⁶ were enumerated by flow cytometry in 10–14-week-old mice. Values shown are the means and standard deviations from N independent analyses (rounded to the nearest 0.5 × 10⁵).

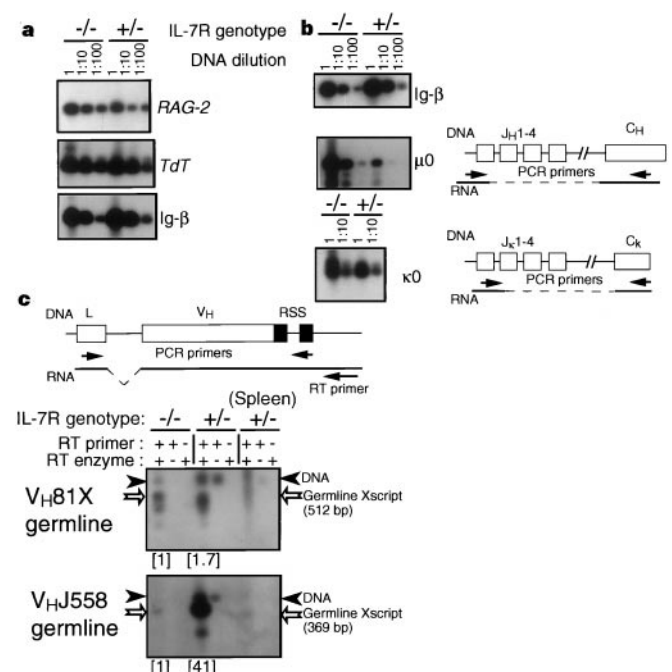


Figure 3 Specific defect in recombination substrates rather than the recombinase machinery. **a**, Normal expression in IL-7Rα^{-/-} pro-B cells of RAG-2 and TdT encoding enzymatic components of the recombinase and of the B-lineage specific Ig-β gene product. Sizes of PCR products were: RAG-2, 542 bp, TdT, 313 bp, Ig-β, 397 bp. Probes were isolated from cloned cDNA by PCR using the same primer pairs. **b**, **c**, Comparison of germline transcripts from different regions of the IgH locus in IL-7Rα^{-/-} and +/- bone marrow. The amount of cDNA in each reaction was normalized to the Ig-β gene product. In **c**, this amount corresponds to '1' in **b**. V_H sterile transcripts were quantified by densitometry of autoradiographs (numbers below tracks represent signal intensity in arbitrary units). The 3' PCR primers correspond to the heptamer-spacer region of the RSS. The 5' PCR primer corresponds to the leader (L) exon of the V_H558 or V_H81X coding sequence, immediately upstream of an intron of ~100 bp. Spurious amplification from contaminating genomic DNA appears as a larger band (filled arrowhead). No products were detected in the absence of primer in reactions containing reverse transcriptase (lanes marked '+ RT enzyme, -RT primer'), ruling out low levels of nonspecific priming. As a further control for specificity, no germline transcripts (open arrows) are detected using RNA from IL-7Rα^{+/+} spleen cells expected to express high levels of rearranged V gene transcripts but no germline transcripts. Similar results were obtained in at least 3 independent experiments.

mice (Table 1), demonstrating that the pool of pro-B cells undergoing $V-(D)J_H$ joining is similar. Further supporting evidence comes from an additional experiment using μ MT mice, which harbour a germline disruption of the membrane exon of the IgM (μ) heavy-chain gene and are therefore incapable of synthesizing membrane-bound μ molecules⁹. Like IL-7R $\alpha^{-/-}$ bone marrow, μ MT bone marrow is deficient in pre-B and B lymphocytes but contains similar numbers of pro-B cells in fractions A–C¹⁰. Nevertheless, recombination of VHJ558 segments is easily detectable in μ MT pro-B cells, in contrast to the comparable IL-7R $\alpha^{-/-}$ fraction (Fig. 2c). Taken together, these results show that the observed decrease in VHJ558 usage in the absence of IL-7R α cannot simply be ascribed to a selective failure in proliferative expansion or a reduction in the pool of cells undergoing $V-(D)J_H$ joining. Instead, it must be due to a reduced frequency of recombination.

This is supported by examination of rearrangements involving other VH families in the murine IgH locus (Fig. 2d) by a similar strategy of PCR using family-specific primers³. Although murine VH families have not yet been precisely mapped, their 5' to 3' order in the IgH locus has been established^{5,6}. We find that VH segments belonging mainly to the most DH-proximal families undergo rearrangement efficiently in purified IL-7R $\alpha^{-/-}$ sIgM⁺ B cells (Fig. 2d). Thus, in the absence of IL-7R α the efficiency of VH recombination decreases with increasing distance from DH/JH. Most VH segments are infrequently rearranged, reducing the recombinational diversity of antibody heavy-chain genes.

In this instance, defective VH recombination cannot readily be explained by faulty recombinase activity. The same $V(D)J$ recombinase mediates all joins^{11,12}, yet the defect we observe is specific to recombination of DH-distal VH segments alone: D–JH as well as VH7183–(D)JH joining occur normally. Moreover, transcripts⁷ from genes such as RAG-2 and TdT, encoding enzymatic components of the recombinase, are expressed at normal levels in IL-7R $\alpha^{-/-}$ bone marrow fractions (Fig. 3a).

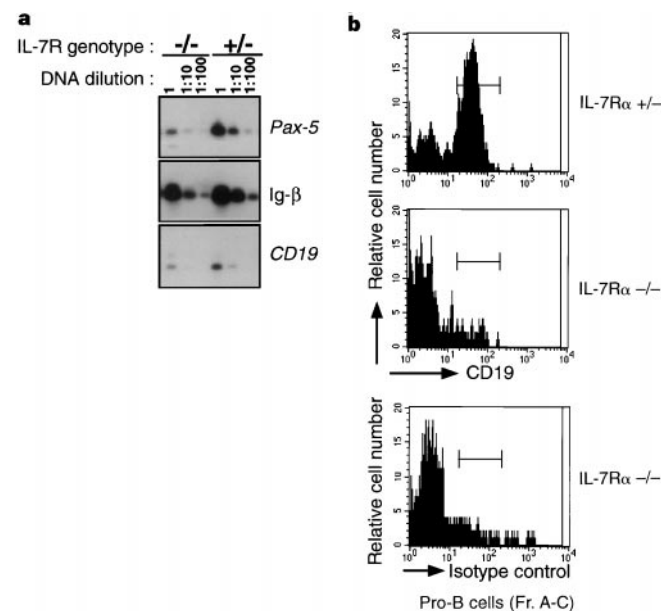


Figure 4 Expression and activity of Pax-5 are diminished in the absence of IL-7R α . **a**, Transcripts encoding Pax-5 are compared in IL-7R $\alpha^{-/-}$ and +/- bone marrow normalized for Ig-β transcripts expressed at the same stages in development. Pax-5 activity is estimated by comparative measurement of transcripts of CD19, whose promoter is a known target gene of Pax-5. Sizes of PCR products were: Pax-5, 440 bp, Ig-β, 397 bp, CD19, 773 bp. Probes were isolated from cloned cDNAs using the same primer pairs. **b**, Surface staining for CD19 in gated B220⁺CD43⁺ bone marrow pro-B cells, with staining of an isotype-matched control antibody below.

For recombinase activity to be effective in causing $V(D)J$ recombination, it must have proper access to its chromosomal substrates¹³. Accessibility is brought about by uncharacterized changes in the chromatin structure of the IgH locus which are marked by tissue- and stage-specific germline transcripts originating in unrearranged regions of the locus^{13–15}. Although it is not known if germline transcription causes or merely reflects substrate accessibility, there is a strong correlation between the two processes^{13–15}. We therefore compared the structure of the IgH locus in bone marrow cells from IL-7R $\alpha^{-/-}$ and +/- mice using PCR after reverse transcription of RNA (RT-PCR) to detect the presence of germline transcripts from three different regions: μ 0 transcripts¹⁶ originate from the JH cluster, and germline transcripts of unrearranged VHJ558 segments¹³ originate from the extreme 5' end of the VH cluster. In addition, we devised an assay to detect germline transcripts from the unrearranged VH81X segment, the member of the 3' VH7183 family most frequently used in rearrangements^{13,17}. These assays are shown in Fig. 3b, c. The amount of RNA in the reactions was normalized to transcripts from the immunoglobulin Ig-β gene which is expressed exclusively in the B-cell lineage from the earliest stages in development⁷. Extensive controls (Fig. 3c) distinguish germline transcripts from unrearranged DNA or transcripts or rearranged V genes. Furthermore, the identity of PCR products corresponding to germline transcripts was verified by DNA sequencing.

μ 0 transcripts and germline transcripts from the unrearranged VH81X segment occur at similar or even higher levels in -/- samples than +/- controls (Fig. 3b, c), as expected from the greater proportion they contain of pro-B cells undergoing IgH rearrangements (Table 1). This indicates that structural changes to the 3' end of the IgH locus proceed normally in IL-7R $\alpha^{-/-}$ bone marrow, consistent with unimpaired D to JH joining (Fig. 1c) and recombination of adjacent VH7183 segments (Fig. 2b, c). In contrast, germline transcripts from unrearranged VHJ558 segments at the extreme 5' end of the locus are undetectable under the same conditions (Fig. 3c) in several independent experiments. The specificity of this defect is emphasized by the observation that κ 0 germline transcription from the unrearranged κ light-chain locus^{15,18}, which also begins in early progenitors is normal in IL-7R $\alpha^{-/-}$ bone marrow (Fig. 3b). Thus, in the absence of IL-7R α there is a selective failure in developmentally regulated changes to the chromatin structure of VH segments located at the 5' end of the IgH locus.

Changes to chromatin structure are initiated by proteins bound to DNA elements termed locus-control regions (LCRs). The Pax-5 gene product binds an LCR located at the 3' end of the IgH locus^{19,20} and is essential for $V-(D)J_H$ joining early in B lymphopoiesis^{21,22}. We therefore examined its expression in mice lacking IL-7R α . Pax-5 expression is reduced in the bone marrow of IL-7R $\alpha^{-/-}$ mice when compared to +/- controls (Fig. 4a), under conditions where there is no decrease in Ig-β transcripts (Fig. 4a) expressed at precisely the same stages during early B-lymphocyte development^{7,23}. That reduced Pax-5 expression results in reduced activity is demonstrated by the reduction in transcripts encoded by CD19 (Fig. 4a), whose promoter is a known target of Pax-5 (ref. 21). Expression of CD19 protein is also diminished in pro-B cells from the -/- mice (Fig. 4b), although this population is numerically equivalent to its +/- counterpart (Table 1) and contains just as many (D)JH and VH7183 rearrangements (Figs 1c, 2c). Taken together, our findings show that Pax-5 function is impaired during early B lymphopoiesis when IL-7R α is absent. As Pax-5 is the only transcription factor known whose absence reduces V to (D)J (but not D to J) recombination, we speculate that it may work downstream of IL-7R signalling.

Our results extend the accessibility model for the regulation of $V(D)J$ recombination^{13,24}. They provide evidence that a cytokine receptor delivers an extrinsic regulatory signal that alters substrate

accessibility to the recombinase, and suggests a mechanism for its action. Our findings indicate that substrate accessibility at the 5' end of the *IgH* locus is controlled separately from the 3' end, with IL-7R α triggering the programmed 3' to 5' shift in V_H segment recombination which normally occurs during B-lymphocyte maturation in adult bone marrow^{8,16,17}. Once recombination of 5' V_H segments has been triggered, their representation in functional IgH molecules may subsequently be increased by selection for binding to surrogate light chains^{17,25}. By enhancing recombinational diversity before selection, IL-7R α ligands may regulate the primary repertoire of antibody specificities generated by V(D)J recombination: a previously unrecognized, antigen-independent influence on the specific immune response. □

Methods

Mice. IL-7R α ^{-/-} mice¹, +/- controls and MT mice⁹ were maintained under pathogen-free conditions and used for experiments at 10–14 weeks of age. Strain background (129 × C57BL/6) of the IL-7R α ^{-/-} and +/- animals was identical whereas μ MT mice were (129 × C57BL/10).

Flow cytometry. Single cell suspensions from femoral bone marrow were analysed by flow cytometry after staining²⁶ with anti-B220 (R-phycoerythrin-coupled, clone RA3-6B2; PharminGen), anti-IgM μ chain (fluorescein-coupled; Southern Biotechnology) and anti-CD43 (biotin-coupled; PharminGen) developed with streptavidin-red⁶⁷⁰ (Gibco-BRL). Also, for 4-colour staining, anti-B220 (allophycocyanin-coupled), anti-BP-1 (fluorescein) and anti-HSA (R-phycoerythrin) were from PharminGen. Cell numbers in the pro-B (fractions A–C of ref. 26), pre-B (fraction D) and B-cell (fractions E, F) compartments were calculated as described¹⁰. Cell sorting was done on a Becton–Dickinson FACStar Plus using the same antibodies.

PCR assays for *IgH* rearrangements. High-molecular-mass genomic DNA was extracted by the proteinase K method from cells purified by fluorescent cell sorting. Two rounds of PCR amplification were done as follows. First round: 96 °C for 3 min, 2 cycles of 94 °C 40 s, 65 °C 45 s, 72 °C, 1 min 45 s, 3 cycles of 94 °C, 63 °C and 72 °C, then 20 cycles of 94 °C, 60 °C and 72 °C for the same times. Second round: 94 °C 3 min, 3 cycles of 94 °C, 67 °C and 72 °C, followed by 24 cycles of 94 °C, 64 °C and 72 °C. (D)_H joins: the sequences of nested 5' primers corresponding to DSP2/DFL16 segments, and nested 3' primers from the intron downstream of J_H4 are as published^{3,27}. V_H to (D)_H joins. The same 3' primers were used. Specific, nested 5' primers corresponding to V_H7183 or V_H558 segments were: V_H7183, 5' CTCGCCATGGACTTCGGG TCAGTTGG-3' (first round), 5'-CAGCTGGTGGAGTCTGGGGGAGGC-3' (second round); V_H558, 5'-ACCATGGGATGGAGCTGKATCWTBC-3' (first round), 5'-GTGARGCCTGGGRCCTCAGTGAAG-3' (second round). The sequence of primers used to detect rearrangements of other V_H families were as published³. PCR products were detected by Southern blot hybridization to specific radiolabelled probes (PCR generated 200-bp (D)_H4 probe for DJ and V_H family rearrangements in Figs 1c and 2d; family-specific probes spanning V_H exon for V_H558 and V_H7183 rearrangements, of 315 and 283 base pairs (bp) respectively). No amplification products were detected in DNA from non-rearranging 3T3 fibroblasts (not shown).

RT-PCR assays and germline transcription. RNA isolated by the guanidinium method (RNAzol, Bio/Gene) was reverse-transcribed using AMV reverse transcriptase (HT Biotechnology) under the manufacturer's recommended conditions. Poly(dT)(12–16) (Pharmacia) was used as the RT primer unless otherwise specified. Serial dilutions of cDNA were PCR-amplified using specific primers and the products detected using appropriate radiolabelled probes. Oligonucleotide primers for amplification of Ig- β and TdT transcripts were as described⁷; those for RAG-2 were: 5' primer, GCTATGTCAGAAGCATCTCT ATTTC; 3' primer, CTTGGCAGGAGTCAAGACTTTCCC; for Pax-5²⁸: 5' primer, CTACAGGCTCCGTGACGCAG; 3' primer, GTCTCGGCCTGTG AAATAGG; and for CD19: 5' primer, GTAGAAGAGGGAGGCAATGTTGT GCTG; 3' primer, AGTTCTCAACAGCCAGAGCCACACTGC. Primers and amplification conditions for μ 0 and κ 0 germline transcripts have been described^{16,29}. For detection of V_H81X and V_H588 germline transcripts, reverse transcription was carried out at 50 °C. The primers were located immediately 3' of the heptamer–nonamer recombination signal sequence (RSS) at the 3' end

of the V_H81X or V_H558 segments, ensuring that only unrearranged V_H transcripts were reverse-transcribed. They are: V_H81X: GTCTCTCCGCG CCCCTGCTGGTCC; and V_H558: CAATCCCAGTGCAGCTTCCTGCTCC. PCR primers corresponding 5' to the leader exon, and 3' to the RSS are as follows. V_H81X: 5' primer, CTCGCCATGGACTTCGGGGCTCAGCTTGG; 3' primer, GTAACCTTTTGCTCATTTGTG; V_H558: 5' primer, ACCATGGGATGG AGCTGKATCWTBC; 3' primer, TGWGGTWSYMWCACTGTG. (K is (G/T); W, (A/T); B, (C/G/T); S, (C/G); and Y, (C/T)). Amplification conditions were: 96 °C 3 min, 3 cycles of 94 °C 40 s, 67 °C 45 s, 72 °C 45 s, then 32 cycles of 96 °C, 65 °C and 72 °C for the same times. PCR products corresponding to germline transcripts were sequenced on an ABI automated sequencer using the following primers: V_H81X, AGAAGACCCTGTACCTGC, and V_H558, AAGGCCACACT-GAC TGCA.

Received 5 August; accepted 27 November 1997.

1. Peschon, J. J. *et al.* Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *J. Exp. Med.* **180**, 1955–1960 (1994).
2. Corcoran, A. E. *et al.* The interleukin-7 receptor α chain transmits distinct signals for proliferation and differentiation during B lymphopoiesis. *EMBO J.* **15**, 1924–1932 (1996).
3. Ehlich, A., Martin, V., Muller, W. & Rajewsky, K. Analysis of the B-cell progenitor compartment at the level of single cells. *Curr. Biol.* **127**–137 **4**, (1994).
4. Chang, Y., Paige, C. J. & Wu, G. E. Enumeration and characterisation of DJ_H structures in mouse fetal liver. *EMBO J.* **11**, 1891–1899 (1992).
5. Rathbun, G. A., Capra, J. D. & Tucker, P. W. Organization of the murine immunoglobulin V_H complex in the inbred strains. *EMBO J.* **6**, 2931–2937 (1987).
6. Honjo, T. & Matsuda, F. in *Immunoglobulin Genes* 2nd edn (eds Honjo, T. & Alt, F. W.) 145–172 (Academic, London, 1995).
7. Li, Y. S., Hayakawa, K. & Hardy, R. R. The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J. Exp. Med.* **178**, 951–960 (1993).
8. Wu, G. & Paige, C. V_H gene family utilisation in colonies derived from B and pre-B cells detected by the RNA colony blot assay. *EMBO J.* **5**, 3475–3481 (1986).
9. Kitamura, D. & Rajewsky, K. Targeted disruption of the μ chain membrane exon causes loss of heavy chain allelic exclusion. *Nature* **356**, 154–156 (1992).
10. Ehlich, A. *et al.* Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development. *Cell* **72**, 695–704 (1993).
11. Shinkai, Y. *et al.* Rag-2 deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867 (1992).
12. Mombaerts, P. *et al.* Rag-1 deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
13. Yancopoulos, G. & Alt, F. W. Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. *Cell* **40**, 271–281 (1985).
14. Alt, F. W., Blackwell, T. K. & Yancopoulos, G. D. Development of the primary antibody repertoire. *Science* **238**, 1079–1087 (1987).
15. Schlissel, M. S. & Baltimore, D. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* **58**, 1001–1007 (1989).
16. Yancopoulos, G. D. *et al.* Preferential utilisation of the most J_H-proximal V_H gene segments in pre-B cell lines. *Nature* **311**, 727–733 (1984).
17. Marshall, A. J., Wu, G. E. & Paige, C. J. Frequency of V_H81X usage during B cell development. *J. Immunol.* **2077**–2084 (1996).
18. Grawunder, U. *et al.* Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional Ig H rearrangement. *Immunity* **3**, 601–608 (1995).
19. Madisen, L. & Groudine, M. Identification of a locus control region in the immunoglobulin heavy chain locus. *Genes Dev.* **8**, 2212–2226 (1994).
20. Michaelson, J. S., Singh, M. & Birshstein, B. K. BSAP: A player at multiple stages of B cell development. *J. Immunol.* **156**, 2349–2351 (1996).
21. Urbanek, P., Wang, Z. Q., Fetka, I., Wagner, E. F. & Busslinger, M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**, 901–912 (1994).
22. Nutt, S. L., Urbanek, P., Rolink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev.* **11**, 476–491 (1997).
23. Adams, B. *et al.* Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.* **6**, 1589–1607 (1992).
24. Yancopoulos, G. D. & Alt, F. W. Regulation of the assembly and expression of variable region genes. *Annu. Rev. Immunol.* **4**, 339–374 (1986).
25. ten Boekel, E., Melchers, F. & Rolink, A. G. Changes in the V_H gene repertoire of developing precursor B lymphocytes in mouse bone marrow mediated by pre-B cell receptor. *Immunity* **7**, 357–368 (1997).
26. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173**, 1213–1225 (1991).
27. ten Boekel, E., Melchers, F. & Rolink, A. The status of Ig loci rearrangements in single cells from different stages of B cell development. *Int. Immunol.* **7**, 1013–1019 (1995).
28. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263–266 (1995).
29. Schlissel, M., Voronova, M. & Baltimore, D. Helix-loop-helix transcription factor E47 activates germline immunoglobulin heavy-chain gene transcription in a pre-T cell line. *Genes Dev.* **5**, 1367–1376 (1991).
30. Kabat, E., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. *Sequences of Proteins of Immunological Interest* (US Department of Health and Human Services, Washington DC, 1991).

Acknowledgements. We thank J. Peschon and J. Sims (Immunex) for the IL-7R α ^{-/-} mouse strain; S. Davies and Neuberger for μ MT bone marrow; C. Rada and C. Napper for DNA sequencing; and D. Fearon, M. Neuberger and K. J. Patel for critically reviewing this manuscript. A.E.C. is a Fellow of the Kay Kendall Leukaemia Research Fund. Work in A.R.V.'s laboratory is funded by the MRC.

Correspondence and requests for materials should be addressed to A.R.V. (e-mail: arv@mrc-lmb.cam.ac.uk).

β -Chemokines are released from HIV-1-specific cytolytic T-cell granules complexed to proteoglycans

Ludwig Wagner*, Otto O Yang*, Eduardo A. Garcia-Zepeda, Yimin Ge, Spyros A. Kalams, Bruce D. Walker, Mark S. Pasternack & Andrew D. Luster

Partners AIDS Research Center and Infectious Disease Units, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts 02129, USA

* These authors contributed equally to this work.

CD8⁺ lymphocytes are believed to be important in host defence against the human immunodeficiency virus (HIV)-1, inhibiting HIV-1 replication through both cytolytic and non-cytolytic pathways^{1–3}. The cytolytic pathway involves calcium-dependent exocytosis of perforin and granzyme proteases, as well as Fas-mediated programmed cell death⁴, whereas the noncytolytic pathway involves the release of chemokines that prevent viral entry⁵. Using granzyme A as a marker of cytolytic granule proteins, and macrophage inflammatory protein (MIP)-1 α and RANTES as markers of HIV-1 inhibitory chemokines, we show that these two very different mediators of viral inhibition are both localized in the cytolytic granules of HIV-1-specific CD8⁺ cytotoxic T lymphocytes (CTL). Following antigen-specific activation, these mediators are secreted together, facilitating both lysis of virion-producing cells and the inhibition of free virus. In addition, RANTES, MIP-1 α and MIP-1 β are secreted by CTL as a macromolecular complex containing sulphated proteoglycans. This association appears to have a functional significance, because heparan sulphate facilitates RANTES inhibition of HIV-1 infection of monocytes.

Proteoglycans serve as the scaffolding for the packaging of the positively charged granzymes and perforin, and are a major constituent of the CTL's cytolytic granules^{6,7}. The interaction of these proteins with proteoglycans within granules is thought to keep these potentially cytotoxic proteins in an inactive intracellular form⁸. The chemokines are basic proteins and bind avidly to negatively charged proteoglycans^{9,10}. It is believed that proteoglycans capture chemokines in the extracellular matrix and on the surface of endothelial cells¹¹, establishing a local concentration gradient from the point source of chemokine secretion. In addition, the interaction of chemokines with proteoglycans is believed to enhance chemokine activity, including activation by interleukin-8 (IL-8) of neutrophils¹² and MIP-1 β -induced T-cell adhesion¹¹.

HIV-1-specific CTL clones secrete significant amounts of RANTES, MIP-1 α , MIP-1 β and glycosaminoglycans following epitope-specific stimulation or anti-CD3 activation (Fig. 1). To determine whether RANTES, MIP-1 α and MIP-1 β are secreted as monomers or as a macromolecular complex bound to glycosaminoglycan-containing proteoglycans, like other effector proteins of the cytotoxic machinery, size-exclusion chromatography was performed on supernatants collected from activated HIV-1-specific CTL clones. Fractions were analysed for granzyme A enzymatic activity as a marker for the granzyme proteases and RANTES, MIP-1 α and MIP-1 β immunoreactivity. As expected, granzyme A enzymatic activity in supernatants of activated HIV-1-specific CTL clones was found in fractions 17–21, corresponding to a complex of relative molecular mass $\sim$ 400K–600K (Fig. 2a, b). Although RANTES, MIP-1 α and MIP-1 β are 8K proteins, they also eluted

predominantly in fractions 16–21, corresponding to a molecular size of 400K–600K (Fig. 2c).

Because the chemokines and granzyme A both bind proteoglycans and elute at the same high relative molecular mass (Fig. 2b, c), we investigated which size fraction contained the secreted CTL sulphated proteoglycans. Thus, HIV-1-specific CTL clones were cultured overnight with ³⁵S-sulphate in order to radiolabel sulphated proteoglycans biosynthetically. Cells were then activated and the supernatant analysed by size-exclusion chromatography (Fig. 2d). A radioactive peak was found in fractions 17–19, again corresponding to M_r 400K–600K, suggesting that there is an association of chemokines and granzyme A with proteoglycans. A second peak, representing unincorporated ³⁵S-sulphate bound to albumin⁷, was found in fractions 25–31.

To characterize further the nature of the association between chemokines and cytotoxic granules, we used dual immunofluorescence confocal microscopy. In unstimulated CTL, granules containing MIP-1 α (red) and granzyme A (green) were uniformly distributed throughout the cytoplasm (Fig. 3). Computer graphic

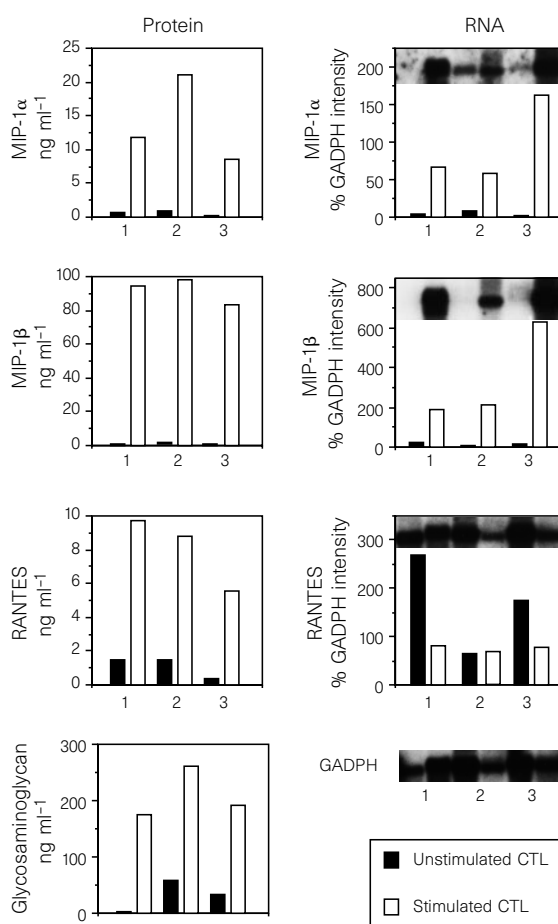


Figure 1 Quantification of chemokine mRNA, protein secretion and sulphated proteoglycan/glycosaminoglycan release from quiescent and activated CTL clones. Eighteen hours after anti-CD3 activation of CTL, supernatant was collected and analysed for MIP-1 α , MIP-1 β and RANTES levels by ELISA and sulphated proteoglycan/glycosaminoglycan by a quantitative dye-binding assay. Total cellular RNA was extracted from the cellular pellet and 10 μ g was analysed by Northern blot analysis with cDNA probes specific for MIP-1 α , MIP-1 β , RANTES, or as a control for RNA loading, glyceraldehyde-3-phosphate dehydrogenase (GADPH). The intensity of the RNA signal was quantified using a phosphor-imager. To normalize for differences in RNA loading, the intensity of the chemokine signal is expressed as a percentage of the GADPH signal.

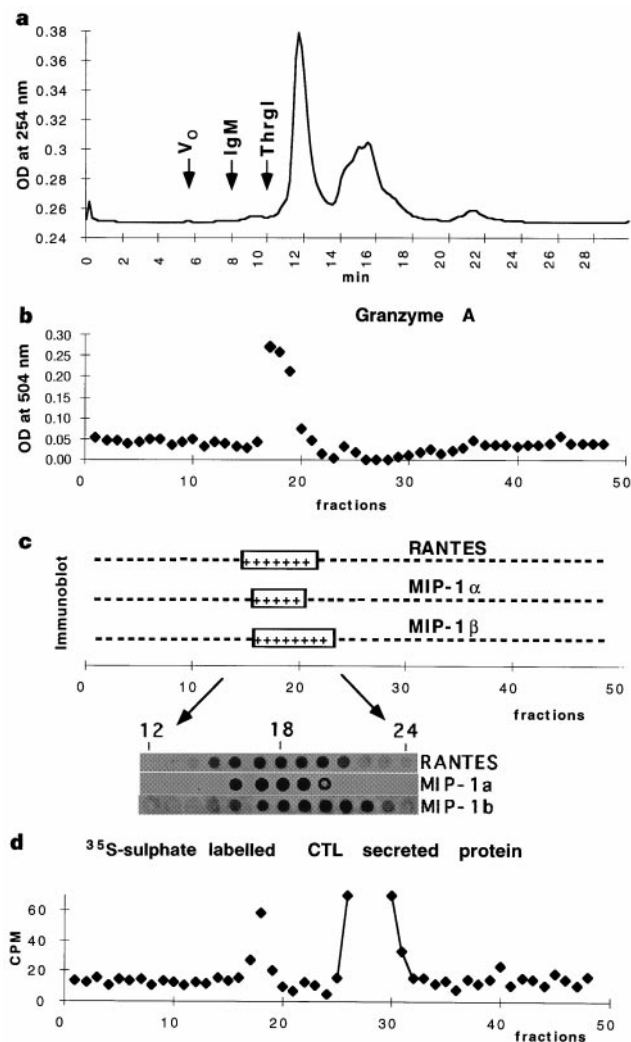


Figure 2 Size-exclusion chromatography of activated CTL supernatant. **a**, Protein elution profile with size marker proteins indicated by arrows (IgM, immunoglobulin M; Thrgl, thyroglobulin). Each fraction was measured for **b**, granzyme A enzyme activity, **c**, MIP-1 α , MIP-1 β and RANTES immunoreactivity, and **d**, ^{35}S -radiolabelled sulphate incorporation, demonstrating that RANTES, MIP-1 α , MIP-1 β and granzyme A are secreted as a large macromolecular complex. This complex has a similar mobility in size-exclusion chromatography to thyroglobulin (400K–600K) and co-elutes with incorporated ^{35}S -sulphate. Shown is a representative experiment for clone 115M21 ($n = 3$); similar results were seen for clone 63D35 ($n = 3$).

Table 1 Production of chemokines by cycloheximide-treated and untreated CTL clones

	Clone 68A62		Clone 15160D75	
Released	No CHX	+CHX	No CHX	+CHX
MIP-1 α	0.4/11.8	0.3/1.4	0.1/48.3	0.0/2.2
MIP-1 β	0.5/140.5	0.4/10.4	0.2/248.0	0.1/10.6
RANTES	1.7/8.3	1.1/4.3	11.1/13.9	3.1/8.8
Intracellular				
MIP-1 α	0.8/1.4	0.6/0.6	0.1/1.9	0.0/0.2
MIP-1 β	0.4/2.8	0.3/0.4	0.1/4.6	0.1/0.5
RANTES	9.9/3.8	3.8/4.7	12.4/6.1	6.0/5.1

CTL clones 68A62 and 15160D75 with and without cycloheximide (CHX) pretreatment were stimulated by CD3 crosslinking for analysis of chemokine production and release. MIP-1 α , MIP-1 β , and RANTES were quantitated in supernatant and cell fractions. Chemokine concentrations are listed as ng per 10^6 cells (unstimulated/stimulated).

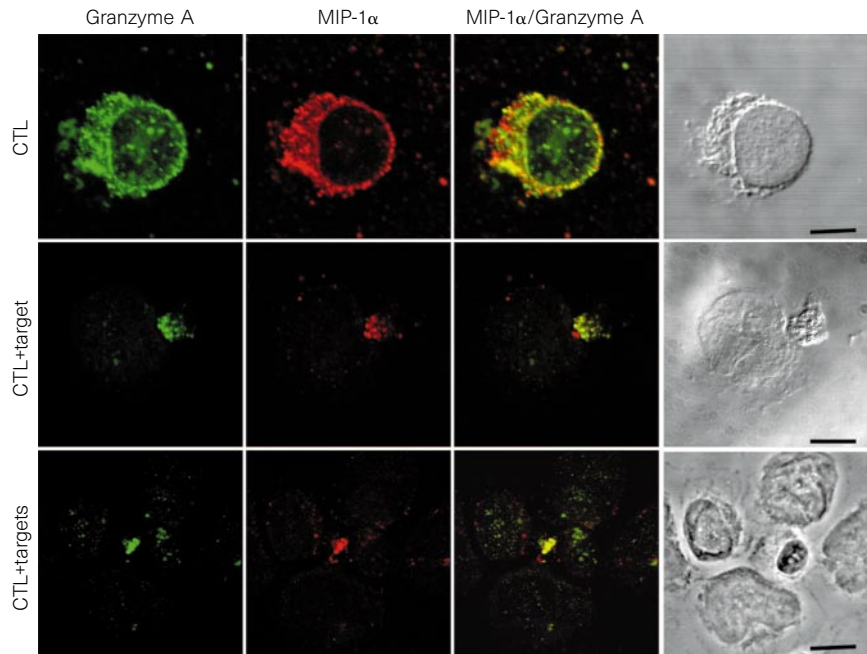


Figure 3 Two-colour immunofluorescence confocal microscopy of an HIV-1-specific CTL clone in the absence of target cells (top row) and in the presence of HIV-1 peptide-loaded MHC compatible target cells (middle and bottom rows). Granzyme A staining is green, MIP-1 α staining is red, and colocalized granzyme A and MIP-1 α staining is yellow. In the top and middle rows, the microscopic images were recorded using Nomarski optics; in the bottom row the microscopic image was recorded using phase-contrast microscopy. 15 min following coincubation of the CTL clone with cognate-peptide-loaded target cells, a polarization of granzyme A granules (green) and MIP-1 α granules (red) within the CTL at the effector–target cell interface is evident. The majority of these granules have colocalized with MIP-1 α and granzyme A (yellow). Scale bars: top, 4.8 μm ; centre and bottom, 8 μm . Shown is a representative of the 3 clones tested ($n = 7$).

analysis demonstrated that MIP-1 α and granzyme A immunofluorescence were strongly co-localized in the cytoplasmic granules of CTLs (yellow in Fig. 3). A minority of the granules contained only MIP-1 α (red) and were localized in the outer rim of cytoplasm in close apposition to the cell membrane. Similarly, there were few cytoplasmic granules that only contained granzyme A immunoreactivity (green) and which were distributed throughout the cytoplasm. Similar results were obtained for RANTES and granzyme A (data not shown). In contrast, stimulation of CTL by incubation with cognate-peptide-loaded target cells induced a reorientation of granules containing granzyme A (green) and MIP-1 α (red) in cytoplasm abutting the area of contact with target cells (Fig. 3). The granules within this polarized region of cytoplasm had markedly augmented co-localization of MIP-1 α and granzyme A (yellow), and were increased in size. This co-localization was also observed for RANTES and granzyme A (data not shown). The intracellular packaging of MIP-1 α and RANTES in the cytolytic T-cell granules bound to sulphated proteoglycans resembles the compartmentalization of two chemokines, platelet factor-4 and platelet basic protein, which are stored in platelet α granules bound to proteoglycans and are released following platelet activation^{13,14}. These findings suggest that storage of chemokines in specialized granules bound to proteoglycans may be a more generalizable phenomenon for effector leukocytes.

To examine the regulation of RANTES, MIP-1 α and MIP-1 β biosynthesis and secretion, chemokine messenger RNA and protein

levels were measured in quiescent and activated CTLs. Analysis of quiescent CTL revealed constitutive RANTES mRNA and intracellular protein expression, and much lower levels of MIP-1 α and MIP-1 β (Fig. 1 and Table 1). Activation increased the amounts of mRNA and intracellular protein of MIP-1 α and MIP-1 β by 5- to 50-fold, while slightly decreasing those of RANTES (Fig. 1 and Table 1). However, activation led to the extracellular release of 5- to 100-fold more of all three chemokines (10–100 ng ml⁻¹) over baseline. Pretreatment of CTLs with cycloheximide before activation, blocking *de novo* protein synthesis, reduced the amount of secreted MIP-1 α and MIP-1 β by 10–20-fold and RANTES by 2-fold (Table 1). These results suggest that the secretion of these chemokines from CTLs following stimulation is the result of the release of both presynthesized stored protein (such as RANTES) and activation-induced *de novo* synthesized protein (for example, MIP-1 α and MIP-1 β).

The finding that CTLs release RANTES, MIP-1 α and MIP-1 β bound to sulphated proteoglycans raises the possibility that this is the form in which these chemokines are seen physiologically, and that this association may modulate their anti-HIV-1 activity. To test this idea, we investigated the effect of heparan sulphate on the ability of recombinant RANTES to inhibit HIV-1 replication in primary lymphoblasts and monocytes. RANTES was suppressive in lymphoblasts in the presence or absence of heparan sulphate (Fig. 4a). In contrast, RANTES was only inhibitory in monocytes in the presence of heparan (Fig. 4b). Heparan sulphate alone had an inhibitory effect on viral replication in lymphoblasts, as shown previously¹⁵, as well as in monocytes (Fig. 4b). The inhibitory effect of heparan alone was additive with that of RANTES in lymphoblasts but synergistic in monocytes. In fact, heparan sulphate at doses that did not demonstrate significant antiviral activity alone (1–10 μ g ml⁻¹) still facilitated RANTES inhibition of the HIV-1 strain JR-CSF in primary monocyte cultures (data not shown). Whereas exogenously added chemokines did not inhibit HIV-1 infection in monocytes, activated CTL supernatant, which contains β -chemokines complexed to proteoglycans, effectively inhibited HIV-1 replication in monocytes (Fig. 4c).

Although other factors released from activated CD8⁺ lymphocytes probably contribute to viral inhibition^{3,16}, the association of chemokines with proteoglycans may be one reason why CD8⁺ T-cell supernatants containing β -chemokines inhibit HIV-1 infection of primary macrophage cultures, whereas recombinant β -chemokines alone generally do not^{16–19}. Recombinant β -chemokines may be ineffective at blocking HIV-1 entry into monocytes because these cells do not, in general, express the cell-surface proteoglycans capable of binding β -chemokines²⁰. However, the expression of proteoglycans in macrophages is a dynamic and highly regulated process that depends on the state of cellular differentiation and activation^{21,22} and this may explain other results indicating that β -chemokines can inhibit envelope-mediated fusion or infection of macrophages^{23,24}. In contrast, lymphocytes constitutively express cell-surface heparan sulphate proteoglycans that bind β -chemokines²⁰, and recombinant β -chemokines inhibit HIV-1 entry into lymphoblasts. Furthermore, digestion of cell-surface heparan sulphate with heparitinase dramatically decreases the ability of β -chemokines to inhibit HIV-1 entry into lymphoblasts²⁰. Our finding that β -chemokines are released from CD8⁺ T cells complexed with proteoglycans suggests that this association can substitute for cell-surface proteoglycans in facilitating chemokine inhibition of HIV-1 entry into monocytes. Thus, the form in which chemokines are secreted and presented to cells is likely to have functional significance. □

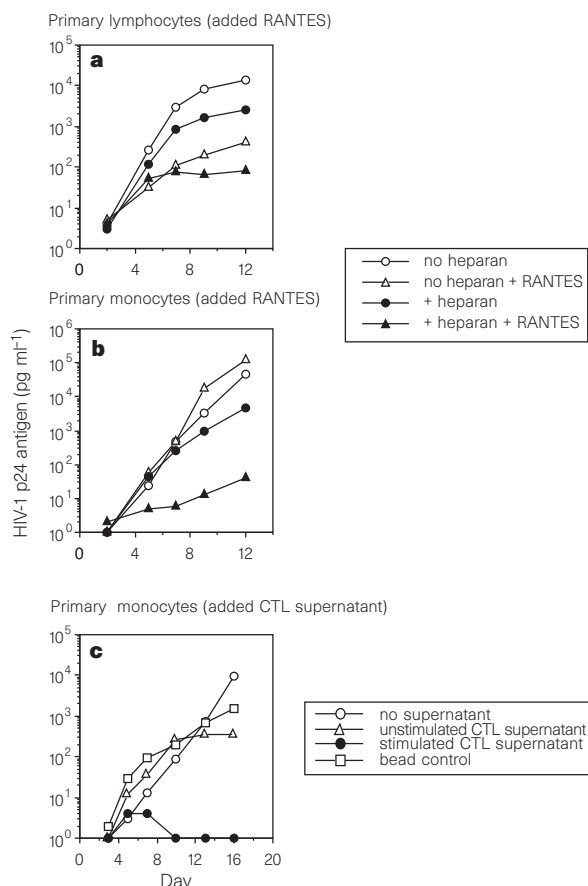


Figure 4 Inhibition of HIV-1 replication by RANTES in the presence or absence of heparan sulphate. Primary lymphoblast (a) and primary monocyte (b, c) cultures from an HIV-1 seronegative donor were acutely infected with the monocytopathic strain HIV-1 JR-CSF and cultured with the indicated combinations of heparan sulfate (100 μ g ml⁻¹) and/or RANTES (100 ng/ml) (a and b) or CTL supernatant (c). Measurement of HIV-1 p24 antigen was performed at 2 to 4 day intervals. This is a representative experiment of $n = 5$ for monocytes and $n = 3$ lymphoblasts.

Methods

Effector cells. HIV-1-specific CTL clones were isolated from infected individuals, characterized, and maintained as described^{25,26} in RPMI 1640 (Sigma) containing 10% FCS and 50 U IL-2 per ml (R10-50). These effector

cells included 115M21 and 15160D75 (HLA B14-restricted, recognizing the envelope Gp41 epitope ERYLKDQQL), 63D35 (HLA A11-restricted, recognizing the HIV-1 reverse transcriptase epitope AIFQSSMTK), 13010A3 (HLA A2-restricted, recognizing the Gag p17 epitope SLYNTVATL), and 161JD27 (HLA B60-restricted, recognizing the Gag epitope IEKDTKEAL). These cells were used at least 2 weeks after restimulation with soluble anti-CD3 antibody 12F6 and irradiated allogeneic feeder cells as described³.

Generation and analysis of activated CTL supernatants. CTL clones were activated using anti-CD3-crosslinking beads (3 beads per cell) in R10-50 as described³. Analysis of chemokine production: CTL clones at 2.5×10^5 cells per ml were activated overnight and the concentrations of MIP-1 α , MIP-1 β , and RANTES in the supernatant were determined using quantitative ELISA assays (R&D Systems) and cellular RNA was isolated using RNA STAT 60 (Tel-Test B, Friendswood, Texas). 10 μ g total cellular RNA was fractionated on a 1.2% agarose/0.7% formaldehyde gel, transferred to GeneScreen (Dupont) and sequentially hybridized with ³²P-radiolabelled cDNA probes. Blots were washed at high stringency ($0.2 \times$ SSC, 55°C) and exposed for 10 days (MIP-1 α , MIP-1 β , RANTES) or 18 h (GADPH). A quantitative dye-binding method (blyscan assay; Accurate Chemical & Scientific, Westbury, NY) was used to determine the concentration of sulphated proteoglycans and glycosaminoglycans in CTL supernatants. Protein synthesis was inhibited by preincubating CTLs with cycloheximide (8 μ g ml⁻¹, a concentration previously determined to yield >90% cell viability and >90% inhibition of protein synthesis) for 1 h at 37°C. CTLs were then activated at 10^6 cells per ml overnight and chemokine measurements determined on the supernatants and on the cellular pellets following lysis in phosphate-buffered saline (PBS), 1% Triton-X-100, and 2 mM diisopropyl fluorophosphate. Size-exclusion analysis of CTL supernatants: CTL clones at 2.5×10^5 cells per ml were activated for 2–4 h and filtered supernatant aliquots were analysed by high-performance size-exclusion chromatography using a Phenomenex Biosep SEC-4000 7.8 $\times$ 300 mm column employing PBS as the mobile phase at 1 ml per min; 500- μ l fractions were collected. Fractions containing exocytosed granzyme A were identified colorimetrically using benzyloxy carbonyl l-lysyl thiobenzylester (BLT) and Ellman's reagent as described²⁷. These fractions were also analysed for the presence of chemokines by immunoblot analysis. Fractionated secreted CTL supernatant was loaded onto nitrocellulose membranes using a multiwell filtration manifold. The membranes were blocked with 5% skim milk and then incubated sequentially with monospecific rabbit anti-human chemokine antiserum (PeproTech) monospecific biotinylated goat anti-rabbit immunoglobulin, an avidin-biotinylated horseradish peroxidase complex (Vector Laboratories), and developed using Renaissance chemiluminescence reagent (NEN Life Science Products). To label CTL proteoglycans biosynthetically, 50 μ Ci of ³⁵S-sulphate was added to 2 ml CTL cultures for 18 h, washed, and activated with anti-CD3 beads. After 3 h, secreted material was analysed by size-exclusion chromatography as above and aliquots of each fraction were counted in a β counter.

Cytospin preparation and staining. The CTL clones 115M21, 13010A3 and 161JD27 were used. The transformed cell line T1, expressing HLA A2 and B60, served as the target cell for the CTL clones 13010A3 and 161JD27. T1 cells (1×10^6) were incubated with the cognate peptide (100 μ g ml⁻¹) and, after 15 min, 3×10^5 specific CTL were added to the target cells and incubated at 37°C. After 10, 15, 20, 30 and 40 min aliquots of 5×10^5 per ml were cytospun onto slides, air-dried for 60 min, fixed in acetone for 5 min, blocked with 3% bovine serum in PBS and then incubated with mouse anti-human MIP-1 α mAb (1:300) (PeproTech) or mouse anti-human RANTES (1:300) (R&D Systems) and rabbit anti-granzyme A²⁸ (1:700) for 90 min. After 2 washes in PBS for 15 min, slides were incubated for 90 min in fluorescein-conjugated donkey anti-rabbit IgG and rhodamine-conjugated donkey anti-mouse IgG (both diluted 1:200) (Jackson ImmunoResearch). Control slides were incubated with only secondary antibodies. After staining, slides were washed in PBS and a drop of 1% n-propyl gallate (Sigma) was used to reduce fluorescence photobleaching. All specimens were examined by confocal imaging using a Leica DM IRBE microscope and the Leica TCS 4D confocal system.

HIV-1 infection of primary lymphoblasts and monocytes. Primary lymphoblasts from a seronegative donor were generated by stimulating Ficoll–Hypaque-purified PBMC with a CD3/8 bispecific monoclonal antibody as previously described³, yielding >95% CD3⁺, CD4⁺ cells by flow cytometric analysis (data not shown). Primary monocytes were adherence-selected by

addition of 5×10^6 PBMC per well to polystyrene 24-well plates for 2 h at 37°C. The wells were washed 3 times and cultured for 2 days in RPMI 1640 with 10% FCS. The remaining adherent cells prepared by this method were <2% CD3⁺ and >85% CD14⁺ by flow cytometric analysis (data not shown). Lymphoblasts were acutely infected by a 4-h incubation with HIV-1 JR-CSF³ at MOI of 10^{-2} at 10 d after stimulation, and plated in a 24-well plate at 5×10^5 cells per well. Monocytes were infected on day 3 after isolation with 5,000 TCID₅₀ of HIV-1 JR-CSF added to each well (estimated to yield a MOI of 10^{-2}), followed by 3 washes. The infected lymphoblast and monocyte cultures were then incubated with 2 ml of R10-50 with or without RANTES (100 ng ml⁻¹) (PeproTech) and/or heparan sulphate (1–100 μ g ml⁻¹) (Sigma) as indicated. At 2–4-day intervals, 1 ml of supernatant was removed and replaced with medium containing RANTES/heparan as appropriate. Supernatant HIV-1 p24 antigen concentrations were determined by quantitative ELISA (Dupont).

Received 9 October; accepted 5 December 1997.

- Walker, C. M., Moody, D. J., Stites, D. P. & Levy, J. A. CD8⁺ lymphocytes can control HIV infection *in vitro* by suppressing virus replication. *Science* **234**, 1563–1566 (1986).
- Tsubota, H., Lord, C. I., Watkins, D. L., Morimoto, C. & Letvin, N. L. A cytotoxic T lymphocyte inhibits acquired immunodeficiency syndrome virus replication in peripheral blood lymphocytes. *J. Exp. Med.* **169**, 1421–1434 (1989).
- Yang, O. O. et al. Suppression of HIV replication by CD8⁺ cells: evidence for HLA class I-restricted triggering of cytolytic and non-cytolytic mechanisms. *J. Virol.* **71**, 3120–3128 (1997).
- Berke, G. The CTL's kiss of death. *Cell* **81**, 9–12 (1995).
- Cocchi, F. et al. Identification of RANTES, MIP-1 α , and MIP-1 β as the major HIV-suppressive factors produced by CD8⁺ T cells. *Science* **270**, 1811–1815 (1996).
- MacDermott, R. P. et al. Proteoglycans in cell-mediated cytotoxicity. Identification, localization, and exocytosis of a chondroitin sulfate proteoglycan from human cloned natural killer cells during target cell lysis. *J. Exp. Med.* **162**, 1771–1787 (1985).
- Stevens, R. S. et al. Co-sedimentation of chondroitin sulfate A glycosaminoglycans and proteoglycans with the cytolytic secretory granules of rat large granular lymphocyte (LGL) tumor cells, and identification of mRNA in normal and transformed LGL that encodes proteoglycans. *J. Immunol.* **139**, 863–868 (1987).
- Tschopp, J. & Masson, D. Inhibition of the lytic activity of perforin (cytolysin) and of late complement components by proteoglycans. *Mol. Immunol.* **24**, 907–913 (1987).
- Witt, D. P. & Lander, A. D. Differential binding of chemokines to glycosaminoglycan subpopulations. *Curr. Biol.* **4**, 394–400 (1994).
- Luster, A. D., Greengard, S. & Leder, P. The IP-10 chemokine binds to a specific cell surface heparan sulfate site shared with platelet factor 4 and inhibits endothelial cell proliferation. *J. Exp. Med.* **182**, 219–231 (1995).
- Tanaka, Y. et al. T-cell adhesion induced by proteoglycan-immobilized cytokine MIP-1 β . *Nature* **361**, 79–82 (1993).
- Webb, L. M. C., Ehrenguber, M. U., Clark-Lewis, I., Baggiolini, M. & Rot, A. Binding to heparan sulfate or heparin enhances neutrophil responses to interleukin 8. *Proc. Natl Acad. Sci. USA* **90**, 7158–7162 (1993).
- Kaplan, K. L., Broekman, M. J., Chernoff, A., Lesznik, G. R. & Drillings, M. Platelet α -granule proteins: Studies on release and subcellular localization. *Blood* **53**, 604–618 (1979).
- Levine, S. P., Knieriem, L. K. & Pager, M. A. Platelet factor 4 and the platelet secreted proteoglycan: Immunologic characterization by crossed immunoelectrophoresis. *Blood* **75**, 902–910 (1990).
- Patel, M. et al. Cell-surface heparan sulfate proteoglycans mediate HIV-1 infection of T-cell lines. *AIDS Res. Human Retro.* **9**, 167–174 (1993).
- Moriuchi, H., Moriuchi, M., Combadiere, C., Murphy, P. M. & Fauci, A. S. CD8⁺ T-cell-derived soluble factor(s), but not β -chemokines RANTES, MIP-1 α , and MIP-1 β , suppress HIV-1 replication in monocyte/macrophage. *Proc. Natl Acad. Sci. USA* **93**, 15341–15345 (1996).
- Dragic, T. et al. HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667–673 (1996).
- Schmidt-Mayerova, H., Sherry, B. & Bukrinsky, M. Chemokines and HIV replication. *Nature* **382**, 767 (1996).
- Simmons, G. et al. Potent inhibition of HIV-1 infectivity in macrophages and lymphocytes by a novel CCR5 antagonist. *Science* **276**, 276–279 (1997).
- Oravecz, T. et al. Regulation of Anti-HIV-1 activity of RANTES by heparan sulfate proteoglycans. *J. Immunol.* **159**, 4587–4592 (1997).
- Uhlén-Hansen, L., Eskeland, T. & Kolset, S. O. Modulation of the expression of chondroitin sulfate proteoglycan in stimulated human monocytes. *J. Biol. Chem.* **264**, 14916–14922 (1989).
- Yeaman, C. & Rapraeger, A. C. Membrane-anchored proteoglycans of mouse macrophages: P388D1 cells express a syndecan-4-like heparan sulfate proteoglycan and a distinct chondroitin sulfate form. *J. Cell. Phys.* **157**, 413–425 (1993).
- Alkhatib, H. et al. CC CKR5: a RANTES, MIP-1 α , MIP-1 β receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955–1958 (1996).
- Verani, A. et al. C-C chemokines released by lipopolysaccharide (LPS)-stimulated human macrophages suppress HIV-1 infection in both macrophages and T cells. *J. Exp. Med.* **185**, 805–816 (1997).
- Walker, B. D. et al. Long-term culture and fine specificity of human cytotoxic T-lymphocyte clones reactive with human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **86**, 9514–9518 (1989).
- Kalams, S. A. et al. Longitudinal analysis of T cell receptor (TCR) gene usage by human immunodeficiency virus 1 envelope-specific cytotoxic T lymphocyte clones reveals a limited TCR repertoire. *J. Exp. Med.* **179**, 1261–1271 (1994).
- Pasternack, M. S. & Eisen, N. H. A novel serine esterase expressed by cytotoxic T lymphocytes. *Nature* **314**, 743–745 (1985).
- Pasternack, M. S., Verret, C. R., Liu, M. A. & Eisen, N. H. Serine esterase in cytolytic T lymphocytes. *Nature* **322**, 740–743 (1986).

Acknowledgements. We thank R. Stevens and M. Norcross for helpful discussions.

Correspondence and requests for materials should be addressed to A.D.L. (e-mail: luster@helix.mgh.harvard.edu).

Carrier protein import into mitochondria mediated by the intermembrane proteins Tim10/Mrs11 and Tim12/Mrs5

Christian Sirrenberg, Maxi Endres, Heike Fölsch, Rosemary A. Stuart, Walter Neupert & Michael Brunner

Institut für Physiologische Chemie der Universität München, Goethestraße 33, 80336 München, Germany

Import of nuclear-encoded precursor proteins into mitochondria and their subsequent sorting into mitochondrial subcompartments is mediated by translocase enzymes in the mitochondrial outer and inner membranes¹⁻³. Precursor proteins carrying amino-terminal targeting signals are translocated into the matrix by the integral inner membrane proteins Tim23 and Tim17 in cooperation with Tim44 and mitochondrial Hsp70 (refs 4-7). We describe here the discovery of a new pathway for the transport of members of the mitochondrial carrier family and other inner membrane proteins that contain internal targeting signals. Two related proteins in the intermembrane space, Tim10/Mrs11 (ref. 8) and Tim12/Mrs5 (ref. 9), interact sequentially with these precursors and facilitate their translocation across the outer membrane, irrespective of the membrane potential. Tim10 and Tim12 are found in a complex with Tim22, which takes over the precursor and mediates its membrane-potential-dependent insertion into the inner membrane. This interaction of Tim10 and Tim12 with the precursors depends on the presence of divalent metal ions. Both proteins contain a zinc-finger-like motif with four cysteines and bind equimolar amounts of zinc ions.

The yeast genes *MRS11/TIM10* and *MRS5/TIM12* encode two related proteins, Tim10 and Tim12 (Fig. 1a) which are essential for the viability of *Saccharomyces cerevisiae*^{8,9}. We constructed the yeast strains Tim10(Gal10) and Tim12(Gal10), in which the *Gal10* promoter was integrated into the chromosome in front of *TIM10* and *TIM12*.

Mitochondria were prepared from Tim10(Gal10) cells, Tim12(Gal10) cells, W334 cells (wild-type) and from Tim22(Gal10) cells³ which were shifted for 15 h to galactose-free medium (indicated by ↓). Tim10↓ mitochondria contained no Tim10 or Tim22 and levels of Tim12 were normal (Fig. 1b); Tim12↓ mitochondria contained neither Tim12 nor Tim22 but normal levels of Tim10. Thus, Tim10 and Tim12 are required for maintaining Tim22. Tim22↓ mitochondria contained Tim10 and Tim12. When these cells were cultured for 35 h in galactose-free medium, the amount of the ADP/ATP carrier AAC was reduced but levels of porin, Tom40, cytochrome *b*₂, Tim23, Tim17, Tim44 and Mge1 were normal (not shown). After 50 h, cells stopped growing. At this stage mitochondrial integrity was severely affected, cytochrome spectra were altered and the precursor of Hsp60 accumulated (results not shown).

Tim10 and Tim12 are located in the intermembrane space^{8,9}. They were associated with the membrane in mitoplasts from wild-type but not in mitoplasts from Tim22↓ mitochondria (Fig. 1c). Apparently, Tim22 is required for the association of Tim10 and Tim12 with the outer face of the inner membrane. The association of Tim10 was salt-sensitive, whereas Tim12 remained membrane-associated even at high ionic strength (results not shown). To test whether Tim10 and Tim12 are in complex with Tim22, mitochondria were solubilized with digitonin and immunoprecipitated with

antibodies against Tim12 and Tim23 and the immunoprecipitates were analysed for Tim components (Fig. 1d). Tim12-specific IgG precipitated Tim12 and, in addition, Tim10 and Tim22; Tim17 and Tim23 did not precipitate. Tim23 antibodies, on the other hand, co-precipitated Tim17 but not Tim10, Tim12 or Tim22. This observation was supported by gel filtration in the presence of digitonin: Tim12 and a fraction of Tim10 co-eluted with Tim22 in a high-molecular-mass form (*M*_r 300K) but were recovered as a smaller particle (70K) when Tim22↓ mitochondria were used (results not shown). Neither Tim10 nor Tim12 co-eluted with the Tim17–Tim23 complex (*M*_r 250K)¹⁰. These results indicate that Tim10 and Tim12 are organized in a complex with Tim22 and this complex is distinct from the Tim17–Tim23 complex. We found that the association of Tim10 with the Tim22 complex is labile, which may reflect a dynamic interaction of functional significance.

Tim10↓ and Tim12↓ mitochondria were defective in import of AAC, PiC (ref. 3) and Mrs3 (ref. 11), members of the carrier family, and in import of Tim22, which contains an internal targeting signal but is not a member of the carrier family (Fig. 2a). Import of these precursors was not affected in Tim23^{fs} mitochondria, which are defective in Tim23. In contrast, preproteins with an N-terminal presequence, in particular pHsp60, were imported into Tim10↓ and Tim12↓ mitochondria but not into Tim23^{fs} mitochondria. Thus, the Tim22 complex appears to be a translocase facilitating import and insertion of inner membrane proteins that are synthesized without a matrix targeting signal; neither Tim10 nor Tim12 are required for import of pHsp60 into the matrix. The accumulation of pHsp60 in cells in which Tim10 and Tim12 were downregulated for

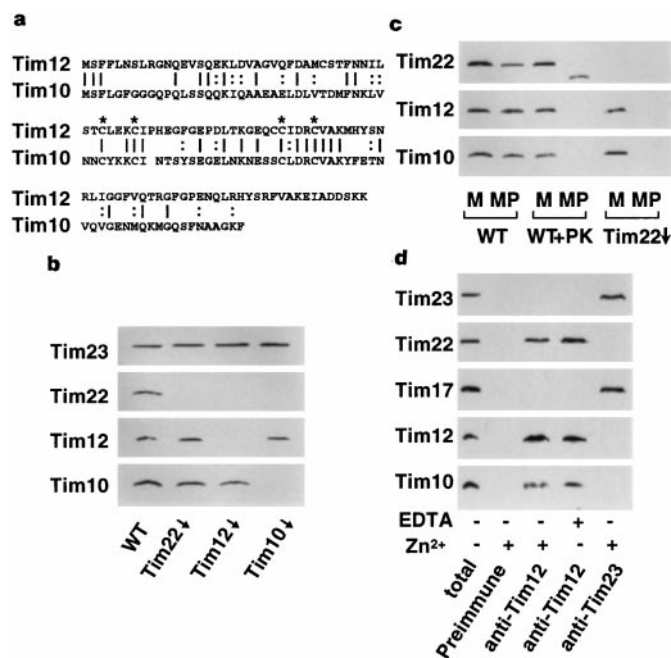


Figure 1 Association of Tim12 and Tim10 with the inner membrane. **a**, Sequence alignment of Tim12 and Tim10. Asterisks indicate conserved cysteines. **b**, Tim10 and Tim12 are required to maintain Tim22. Mitochondria were analysed by immunoblotting. **c**, Membrane association of Tim10 and Tim12 requires Tim22. Mitochondria (M) and mitoplasts (MP) from wild-type and Tim22↓ were analysed with the antibodies indicated. PK, 100 μ g ml⁻¹ proteinase K. **d**, Interaction of Tim12 and Tim10 with Tim22. Mitochondria were solubilized with 1% digitonin and immunoprecipitated. EDTA: 10 mM EDTA/10 mM *o*-Phe; Zn²⁺: 1 mM zinc acetate. Immunoprecipitates were analysed by western blotting.

prolonged times^{8,9} is likely to be a secondary effect and probably does not indicate that Tim10 and Tim12 participate in the import of Hsp60.

To determine whether Tim10 and Tim12 are in contact with AAC during import, we used the crosslinking reagent *m*-maleimido-benzoyl-*N*-hydroxysuccinimide ester (MBS) to crosslink AAC and found one major adduct of 50K (Fig. 2b, upper panel). This crosslinked product was precipitated with antibodies against Tim10. The PiC precursor was also crosslinked to Tim10 during import, but endogenous carrier proteins were not (results not shown). Thus, Tim10 was in contact with carrier proteins during

import. Crosslinking by MBS of AAC and Tim22 was less efficient and was only detected by immunoprecipitation³. MBS did not crosslink AAC with Tim12.

Using 1,4-di(3'-2'-pyridyldithio-(propionamido)butane) (DPDPB) as crosslinker, several adducts with the AAC precursor were generated (Fig. 2b, lower panel). One (*M_r* 42K) was precipitated with antibodies against Tim12, indicating that Tim12 interacts with AAC during import. An adduct of 35K could represent crosslinked Tim5, a component required for transfer of AAC from receptors on the outer membrane to the import pore¹².

To determine the stage along the import pathway at which Tim10 and Tim12 function, we incubated mitochondria with AAC in an import reaction (Fig. 2c). Translocation of AAC across the outer membrane (stage 3)^{3,13} was then assessed by treatment of the mitochondria with proteinase K. To monitor insertion of AAC, mitoplasts were generated and treated with proteinase K. In the presence of a membrane potential, $\Delta\psi$, AAC was imported into the wild-type mitochondria and inserted into the inner membrane, whereas it accumulated at stage 3 in the absence of $\Delta\psi$. When AAC was incubated with Tim10⁻ mitochondria, the preprotein was degraded by proteinase K, indicating that it did not reach stage 3. In contrast, AAC was translocated across the outer membrane in Tim12⁻ and Tim22⁻ mitochondria, but was not inserted into the inner membrane. Thus, Tim10 is required for the initial translocation of AAC across the outer membrane; Tim12 and Tim22 are required for subsequent steps.

To analyse the interaction of Tim10 with precursors, mitochondria were incubated with AAC and crosslinked with MBS (Fig. 2d, upper panel). Generation of the 40K crosslinked product of AAC with Tim10 was efficient in wild-type mitochondria in the presence of $\Delta\psi$: the yield of crosslinked AAC–Tim10 was increased when $\Delta\psi$ was dissipated. This crosslinked complex was absent when AAC was incubated with Tim10⁻ mitochondria but was present in Tim12⁻ and Tim22⁻ mitochondria. Thus, neither Tim12 and Tim22 nor $\Delta\psi$ are required for the interaction of Tim10 with AAC.

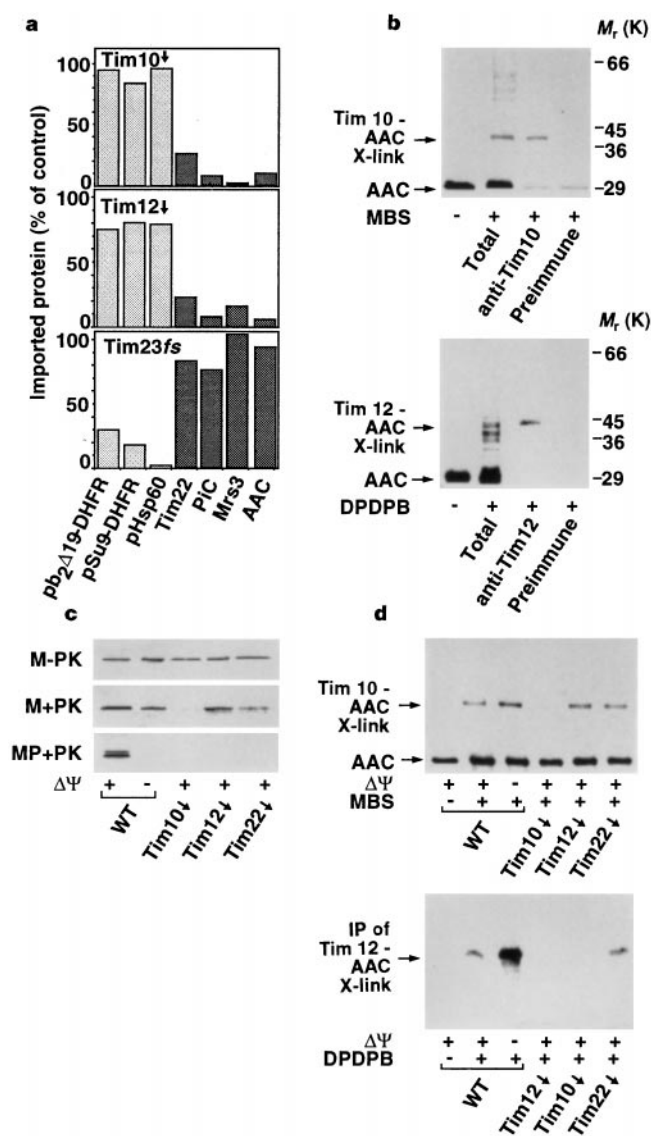


Figure 2 Tim10 and Tim12 facilitate import of carrier proteins and of Tim22. **a**, Import into Tim10⁺, Tim12⁻ and Tim23fs mitochondria³ of matrix-generated pb₂Δ19(1-167)DHFR, pSu9(1-69)DHFR and pHsp60 and inner membrane precursors AAC, PiC, Mrs3 (ref. 17) and Tim22. Imported protein was quantified³ and expressed as per cent import into wild-type (WT) mitochondria. **b**, Tim10 and Tim12 contact AAC during import. AAC was imported at 8°C with 200 μM MBS (top) or at 25°C with 200 μM DPDPB (bottom). Crosslinked products and immunoprecipitates are shown. **c**, Tim10 acts upstream of Tim12 and Tim22. AAC was imported into mitochondria as indicated. M+PK; (5 μg ml⁻¹ PK); MP+PK, mitoplasts treated with 200 μg ml⁻¹ PK. **d**, Sequential interaction with AAC of Tim10, Tim12 and Tim22. Top, AAC was imported (30 min, 8°C) with 200 μM MBS; total crosslinks are shown. Bottom, import (15 min, 25°C) with 200 μM DPDPB; immunoprecipitates with antibodies against Tim12 are shown.

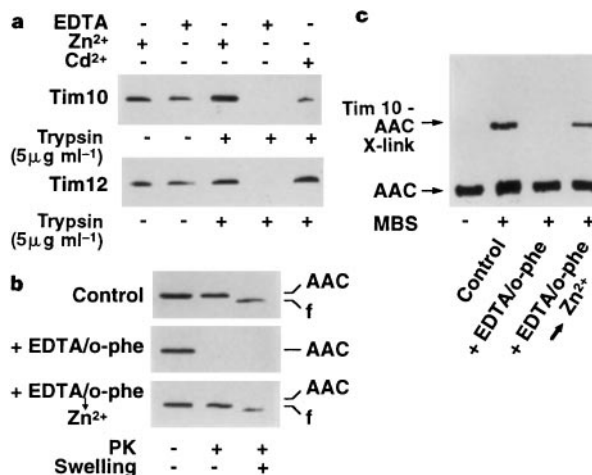


Figure 3 Requirement for divalent cations. **a**, Zn²⁺ and Cd²⁺ stabilize the structure of Tim10 and 12. Trypsin treatment of mitoplasts: EDTA, 10 mM EDTA; Zn²⁺, 1 mM Zn acetate; Cd²⁺, 0.1 mM CdCl₂. **b**, Import of AAC is dependent on divalent cations. Import of AAC (10 min, 12°C) under different conditions; control, import buffer²¹; EDTA/o-phe, 10 mM EDTA, 2 mM *o*-phe; EDTA/o-phe → Zn²⁺, preincubation with EDTA/o-phe followed by resuspension of mitochondria and import with 0.1 mM Zn²⁺. Samples were divided into three aliquots: PK, 200 μg ml⁻¹ PK; swelling, mitoplasts were prepared; f, protease-resistant fragment indicative of inserted AAC. **c**, Zn²⁺-dependent interaction of AAC with Tim10. Mitochondria were pretreated with valinomycin, incubated with AAC in three parallel reactions as in **b** and crosslinked with MBS.

The interaction of Tim12 with AAC precursor was analysed by crosslinking with DPDPB and immunoprecipitation of the radiolabelled AAC-specific adduct by antibodies raised against Tim12 (Fig. 2d, lower panel). The specific adduct was generated in wild-type mitochondria in the presence of $\Delta\psi$. Its yield increased significantly when AAC was imported in the absence of $\Delta\psi$, indicating that AAC accumulated at Tim12. No crosslinking was detected with Tim12 Δ mitochondria. The AAC–Tim12 adduct was also not detected in Tim10 Δ mitochondria but was present in Tim22 Δ mitochondria. Apparently, Tim12 acts downstream of Tim10 but upstream of Tim22 in the import pathway.

In summary, Tim10 and Tim12 have a function early in the import pathway. Tim10 mediates the initial translocation of AAC across the outer membrane. In the absence of $\Delta\psi$, AAC accumulates at stage 3, when it is in contact with Tim10 and Tim12. As this is a productive translocation intermediate¹³, Tim10 and Tim12 keep the hydrophobic preprotein competent for $\Delta\psi$ -dependent membrane insertion by Tim22.

Tim10 and Tim12 contain four cysteine residues which might constitute a metal-binding site, so we tested whether metal ions could stabilize the conformation of these proteins. Mitoplasts were incubated with EDTA or Zn^{2+} or, after EDTA treatment, with Cd^{2+} ; Tim10 and Tim12 were recovered with the mitoplasts, indicating that divalent cations do not affect their association with Tim22 (Fig. 1d). Treatment of mitoplasts with trypsin revealed that Tim10 was degraded by $150 \mu\text{g ml}^{-1}$ trypsin in the presence of EDTA but not in the presence of Zn^{2+} or Cd^{2+} (Fig. 3a). Likewise, Tim12 was degraded by $5 \mu\text{g ml}^{-1}$ trypsin in the presence of EDTA but not in the presence of Zn^{2+} or Cd^{2+} . Thus, the conformation of Tim10 and Tim12 is influenced by divalent metal ions, a characteristic feature of zinc-finger domains¹⁴. In the presence of Zn^{2+} or Cd^{2+} , which can

substitute for Zn^{2+} in Cys₄ zinc-finger motifs¹⁵, the proteins are folded and resistant to protease but appear to be loosely folded in the absence of divalent metal ions.

We investigated whether import of AAC depends on divalent cations. Mitochondria were incubated with AAC in the absence or presence of the metal-chelating agents EDTA and *o*-phenanthroline (EDTA/*o*-phe) (Fig. 3b). With EDTA/*o*-phe, the precursor associated with mitochondria but its import was blocked. Import of preproteins into the matrix was not inhibited by EDTA/*o*-phe¹⁶. The arrested AAC was imported into the inner membrane when the mitochondria were reisolated in the presence of Zn^{2+} (Fig. 3b), indicating that it was a productive translocation intermediate. Therefore, divalent cations are required for AAC import.

To investigate whether this dependence on divalent cations reflects the action of Tim10 and Tim12, we incubated radiolabelled AAC with mitochondria in the absence of $\Delta\psi$ and crosslinked them with MBS: AAC was crosslinked to Tim10 in the presence of divalent cations (Fig. 3c), but not in the presence of EDTA/*o*-phe. When mitochondria were pretreated with EDTA/*o*-phe and reisolated in import buffer containing Zn^{2+} or Cd^{2+} , however, AAC was crosslinked to Tim10. Similarly, AAC was crosslinked by DPDPB to Tim12 in the presence of divalent cations but not in their absence (results not shown). Thus, the interaction of Tim10 and Tim12 with AAC during its import into the inner membrane depends on metal ions.

We constructed chimaeric proteins consisting of Tim10 or Tim12 and maltose-binding protein (MBP) (Fig. 4a) and examined them by atomic absorption spectroscopy (AAS). The preparations contained 0.92 mol Zn^{2+} per mol MBP–Tim10 and 0.99 mol Zn^{2+} per mol MBP–Tim12 (Fig. 4b, left panel). Control preparations of MBP–Lac α contained only background concentrations of Zn^{2+} . Treatment of MBP–Tim12 with EDTA reduced the Zn^{2+} content and readdition of Zn^{2+} or Cd^{2+} efficiently restored its metal content (Fig. 4b, right panel). Treatment of MBP–Tim12 with *N*-ethylmaleimide (NEM) lowered Zn^{2+} to almost background levels, suggesting that sulphhydryl groups are involved in Zn^{2+} binding. Corresponding results were obtained with MBP–Tim10 (not shown). We conclude that Tim10 and Tim12 are specialized proteins of the intermembrane space that mediate the transfer from the outer to the inner membrane of certain hydrophobic membrane proteins that are synthesized without a presequence.

MRS5/TIM12 and MRS11/TIM10 were previously identified together with a number of other MRS genes as suppressors of a defect in mitochondrial RNA splicing¹⁷. As the proteins they encode are located in the intermembrane space, they were not considered to be directly involved with RNA splicing in the matrix^{8,9}. In view of their direct role in protein import, we propose that these proteins be named Tim12 and Tim10 (ref. 18).

The family of mitochondrial carrier proteins in yeast comprises more than 30 members which all contain internal import signals that have yet to be identified. Carrier proteins are composed of three repeat elements, each of which contains a matrix-exposed sequence motif (a 'carrier signature'^{19,20}). A similar motif is also present in Tim22 (ref. 3). The pattern of charged and hydrophobic amino-acid residues in this motif is complementary to a pattern of charged and hydrophobic residues present in the putative zinc-fingers of Tim10 and of Tim12, but in an antiparallel orientation (Fig. 4c), which might constitute a structural basis for the recognition of precursors by Tim10 and Tim12.

Tim10 and Tim12 cooperate in a complex with Tim22. This array of proteins facilitates import of membrane proteins that do not carry a presequence; they are not required for the import of preproteins into the matrix. Our results define a new import pathway of proteins into mitochondria (Fig. 4d): the precursors pass through the TOM complex in the outer membrane by virtue of their interaction with Tim10—this component is a receptor in the intermembrane space that first binds the preprotein and then

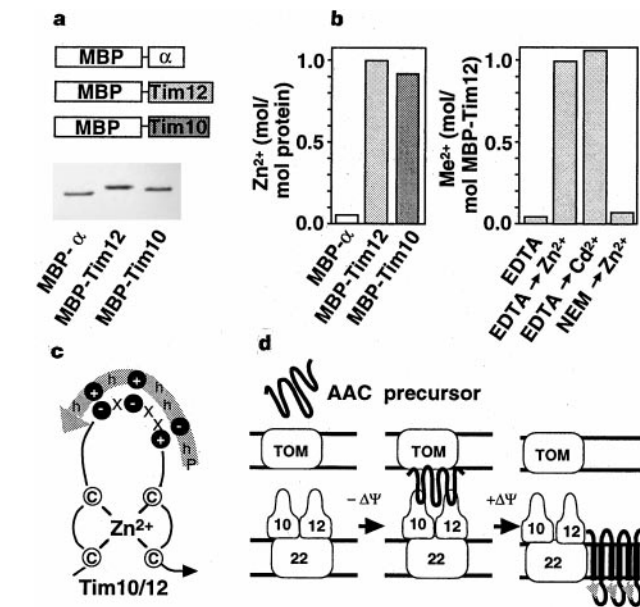


Figure 4 Zn^{2+} content of Tim12 and Tim10 fusion proteins. **a**, Outline and SDS-PAGE of maltose-binding protein (MBP) fused with Tim12, Tim10 or the α -fragment of β -galactosidase (Lac- α). **b**, MBP-Tim12 and MBP-Tim10 bind Zn^{2+} . Left, fusion proteins (1 mg ml^{-1}) were investigated by atomic absorption spectroscopy (AAS) for determination of Zn^{2+} . Right, pretreatment of MBP-Tim12 with EDTA (10 mM EDTA/10 mM *o*-phe), EDTA- Zn^{2+} (readdition of 1 mM zinc), EDTA- Cd^{2+} (readdition of 0.1 mM CdCl_2), NEM- Zn^{2+} (2 mM NEM for 60 min at 0°C , then addition of 1 mM zinc acetate). Samples were reisolated without Zn^{2+} . **c**, Hypothetical interaction of the putative zinc-fingers of Tim10 and Tim12 with carrier signature motifs. Charged (+, -) and hydrophobic (h) residues are indicated. **d**, Model for sequential action of Tim10, Tim12 and Tim22 in import of a carrier. Carrier signature motifs are indicated by arrows.

transfers it to Tim12, which passes it on to Tim22; Tim22 then mediates its $\Delta\psi$ -dependent insertion into the inner membrane. Only at this late stage is $\Delta\psi$ required for the import process. This pathway contrasts with the Tim23–Tim17-mediated pathway that facilitates import of preproteins containing N-terminal targeting signals: here, $\Delta\psi$ is required in an early step to transfer the presequence across both outer and inner membranes to the mitochondrial protein Hsp70, which then coordinates the translocation of the preprotein across both membranes. □

Methods

Subcloning and expression of Tim10 and Tim12. TIM10 and TIM12 DNA were amplified from genomic *S. cerevisiae* DNA by polymerase chain reaction (PCR) and subcloned into pGEM4 and pMAL-cRI. *E. coli* XL-blue cells, transformed with pMAL-cRI-Tim10 and pMAL-cRI-Tim12, were grown at 25 °C to $A_{600} = 0.5$. Zinc acetate (1 mM) was added 15 min before IPTG was added. After 7 h, cells were lysed by sonication in 15 ml buffer A (20 mM HEPES/KOH, pH 7.4, 200 mM NaCl, 10 mM β -mercaptoethanol) containing 1 mM zinc acetate and 1 mg ml⁻¹ lysozyme. The supernatant of a clarifying spin was applied to an amylose column (12 ml) which was then washed with 50 ml buffer A without zinc acetate.

Construction of replacement cassettes. A BamHI–HindIII fragment encoding Tim10 (or Tim12) was subcloned into Yep51mod³, yielding Yep51–Tim10 (or Yep51–Tim12). PCR fragments corresponding to the 5' untranslated region of Tim10 (–297 to –43) and Tim12 (–490 to –136) were cloned into Yep51–Tim10 and Yep51–Tim12, respectively, replacing the 2 μ origin on the HpaI–HindIII fragment. The resulting 'replacement cassette' plasmids³ were linearized with HindIII and used to transform the haploid yeast strain W334. Leu⁺ colonies were analysed.

Received 9 October; accepted 25 November 1997

1. Brunner, M., Schneider, H. C., Lill, R. & Neupert, W. Dissection of protein translocation across the mitochondrial outer and inner membranes. *Cold Spring Harbor Symp. LX*, 619–627 (1995).
2. Kübrich, M., Dietmeier, K. & Pfanner, N. Genetic and biochemical dissection of the mitochondrial protein-important machinery. *Curr. Genet.* **27**, 393–403 (1995).
3. Sirrenberg, C., Bauer, M. F., Guiard, B., Neupert, W. & Brunner, M. Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22. *Nature* **384**, 582–585 (1996).
4. Kronidou, N. G. et al. Dynamic interaction between Isp45 and mitochondrial hsp70 in the protein import system of yeast mitochondrial inner membrane. *Proc. Natl Acad. Sci. USA* **91**, 12818–12822 (1994).
5. Rassow, J. et al. Mitochondrial protein import: biochemical and genetic evidence for interaction of matrix hsp70 and the inner membrane protein MIM44. *J. Cell Biol.* **127**, 1547–1556 (1994).
6. Schneider, H.-C. et al. Mitochondrial Hsp70/MIM44 complex facilitates import. *Nature* **371**, 768–774 (1994).
7. Schneider, H.-C., Westermann, B., Neupert, W. & Brunner, M. The nucleotide exchange factor MGE exerts a key function in the ATP-dependent cycle of mt-Hsp70-Tim44 interaction driving mitochondrial protein import. *EMBO J.* **15**, 5796–5803.
8. Jarosch, E., Rödel, G. & Schwenen, R. J. A soluble 12-kDa protein of the mitochondrial intermembrane space, Mrs11p, is essential for mitochondrial biogenesis. *Mol. Gen. Genet.* **255**, 157–165.
9. Jarosch, E. et al. Mrs5p, an essential protein of the mitochondrial intermembrane space, affects protein import into yeast mitochondria. *J. Biol. Chem.* **271**, 17219–17225 (1996).
10. Bauer, M. F., Sirrenberg, C., Neupert, W. & Brunner, M. Role of Tim23 as voltage sensor and presequence receptor in protein import into mitochondria. *Cell* **87**, 33–41 (1996).
11. Wiesenberger, G., Link, T. A., von Ahsen, U., Walldherr, M. & Schwenen, R. J. MRS3 and MRS4, two suppressors of mtRNA splicing defects in yeast, are new members of the mitochondrial carrier family. *J. Mol. Biol.* **217**, 23–37 (1991).
12. Dietmeier, K. et al. Tom5 functionally links mitochondrial preprotein receptors to the general import pore. *Nature* **388**, 195–200 (1997).
13. Pfanner, N. & Neupert, W. Distinct steps in the import of ADP/ATP carrier into mitochondria. *J. Biol. Chem.* **262**, 7528–7536 (1987).
14. Berg, J. M. & Shi, J. The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **271**, 1081–1085 (1996).
15. Krizek, B. A., Merkle, D. L. & Berg, J. M. Ligand variation and metal ion binding in zinc finger peptides. *Inorg. Chem.* **32**, 937–940 (1994).
16. Schmidt, B., Wachter, E., Sebald, W. & Neupert, W. Processing peptidase of *Neurospora* mitochondria. Two step cleavage of imported ATPase subunit 9. *Eur. J. Biochem.* **144**, 581–588 (1984).
17. Walldherr, M. et al. A multitude of suppressors of group II intron-splicing defects in yeast. *Curr. Genet.* **24**, 301–306 (1993).
18. Pfanner, N. et al. Uniform nomenclature for the protein transport machinery of the mitochondrial membranes. *Trends Biochem. Sci.* **21**, 51–52 (1996).
19. Saraste, M. & Walker, J. E. Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. *FEBS Lett.* **144**, 250–254 (1982).
20. Palmieri, F. Mitochondrial carrier proteins. *FEBS Lett.* **346**, 48–54 (1994).
21. Fölsch, H., Guiard, B., Neupert, W. & Stuart, R. A. Internal targeting signal of the BCS1 protein: a novel mechanism of import into mitochondria. *EMBO J.* **15**, 479–487 (1995).

Acknowledgements. We thank M. Zenk for determination of zinc by AAS. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the Human Frontiers of Science Program, the Fonds der Chemischen Industrie and the Münchener Medizinische Wochenschrift (to R.A.S. and to M.B.).

Correspondence and requests for materials should be addressed to W.N. (e-mail: neupert@bio.med.uni-muenchen.de).

Group II intron splicing *in vivo* by first-step hydrolysis

Mircea Podar*, Vi T. Chu†, Anna Marie Pyle† & Philip S. Perlman*

* Department of Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9148, USA

† Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA

Group I, group II and spliceosomal introns splice by two sequential transesterification reactions¹. For both spliceosomal and group II introns, the first-step reaction occurs by nucleophilic attack on the 5' splice junction by the 2' hydroxyl of an internal adenosine, forming a 2'–5' phosphodiester branch in the intron. The second reaction joins the two exons with a 3'–5' phosphodiester bond and releases intron lariats. *In vitro*, group II introns can self-splice by an efficient alternative pathway in which the first-step reaction occurs by hydrolysis. The resulting linear splicing intermediate participates in normal second-step reactions, forming spliced exon and linear intron RNAs^{2,3}. Here we show that the group II intron first-step hydrolysis reaction occurs *in vivo* in place of transesterification in the mitochondria of yeast strains containing branch-site mutations. As expected, the mutations block branching, but surprisingly still allow accurate splicing. This hydrolysis pathway may have been a step in the evolution of splicing mechanisms.

The branch nucleotide of group II introns is usually an unpaired adenosine located in the 3'-terminal secondary structure, domain 6 (D6)⁴. Five new intron variants with mutations affecting the branch nucleotide (A880) of domain 6 of intron 5 γ of the COXI gene of yeast mitochondrial DNA (aI5 γ) (see Fig. 1b) have been analysed *in vitro* for their effects on self-splicing⁵ (V.T.C., M.P., P.S.P. and A.M.P., manuscript in preparation). Two of these, G and U, were inhibited ~100-fold for branching, but branching was undetectable for the other mutants (C, Δ and P (paired)). Despite these radical branching defects, all five mutants remained highly reactive for self-splicing, using hydrolysis as the first reaction.

We transformed these mutations into otherwise wild-type mtDNAs, and analysed the resulting strains for splicing of the mutated introns. Surprisingly, none of the mutations inhibited splicing enough to block respiratory growth in glycerol cultures at 30 °C, although all grew more slowly, and two mutants, Δ and P, were cold sensitive for glycerol growth at 18 °C (Fig. 1c). RNA blots from cells grown in 30 °C raffinose cultures showed that each strain has a detectable level of mRNA spliced for aI5 γ (Fig. 2a, lanes 3–6, 8), ranging from about 23% of the control level for mutant G (lane 3) to about 2% for mutant P (lane 9). The two cold-sensitive mutants had substantially less splicing in cells grown at 18 °C (Fig. 2a, lanes 7 and 9) than in cells grown at 30 °C (lanes 6 and 8). The levels of spliced mRNA correlate inversely with the doubling times in glycerol medium. Reverse transcriptase polymerase chain reaction (RT-PCR) showed that all of the mutants splice accurately at 30 °C (not shown).

A characteristic feature of wild-type group II intron splicing *in vivo* is that the excised intron lariats are substantially resistant to degradation, probably because of the branch structure^{6,7} (Fig. 2b, lanes 1, 2). Yet these five mutants accumulate none or barely detectable levels of intron RNA (Fig. 2b, lanes 3–6, 8). Thus these branch-point mutations alter splicing in a way that causes the excised intron RNA to be degraded. In wild-type yeast mitochondria, group I introns are excised as linear RNAs that are efficiently degraded^{8–10}. The low levels of excised aI5 γ RNAs in these D6 mutants can be explained by the hypothesis that the

mutant strains carry out the first step of splicing by hydrolysis, so that the resulting linear excised intron RNAs are degraded.

Further characterization of the *in vivo* splicing products supports that interpretation. Polyacrylamide gel analysis shows that the intron RNA from the wild-type strain is mostly lariat, with some migrating like linear intron (or broken lariat) (Fig. 2c, lanes 1, 2; note that wild-type intervening sequence (IVS) RNA is a doublet in the agarose gel of Fig. 2b (lanes 1, 2, 12, 13)). There is very little excised intron RNA in mutants G and U (Fig. 2b, c, lanes 3, 4); darker exposures show that most is linear but some is branched. Mutants C, P and Δ , however, have no detectable lariat; only a trace of linear intron RNA is detectable after long exposure (data not shown).

We found previously that mutant strains with second-step defects exhibit the IVS-E2 RNA intermediate *in vivo*^{11,12}. Presumably this RNA accumulates because it contains the processed 3' end of the *COXI* mRNA needed for mRNA stability¹³. IVS-E2 RNA is barely detectable in the wild-type control but is evident in RNA from the mutants (Fig. 2a–c). We have defined the position and properties of IVS-E2 RNA using RNA from a previously reported mutant, J(56)5 Δ C (here called J4), which has altered spacing between domains 5 and 6 (ref. 12). The IVS-E2 RNA from mutant J4 is a doublet on the agarose gel (Fig. 2a, b, lane 13). Some of the IVS-E2 RNA of mutant J4 is lariat (Fig. 2c, lane 13), and the rest migrates like linear (or broken lariat) IVS-E2 RNA. The branchpoint mutants lack detectable lariat IVS-E2 RNA, and the IVS-E2 RNA present comigrates with the linear IVS-E2 RNA from the J4 mutant on both gel systems (Fig. 2b, c, lanes 3–6, 8).

A reverse-transcriptase primer extension experiment using IVS-

E2 RNA isolated from polyacrylamide gels (Fig. 3a) shows that the D6 mutants have linear (not broken lariat) IVS-E2 RNA. IVS-E2 RNA was analysed from wild-type and mutant J4 (as controls) and the C, Δ and P mutants. Using the upstream intron primer (open arrow), strong stops are detected in each sample at the position of the first intron nucleotide (Fig. 3b, lanes 1, 4–6, 8), showing that the IVS-E2 RNA results from accurate cleavage at the 5' splice junction. Using the downstream exon primer (filled arrow), the IVS-E2 RNA from the control strains results in a strong stop caused by the branch site (Fig. 3c, lanes 1, 8), whereas the RNAs from the D6 mutants are not branched by this test (lanes 4–6). Importantly, these lanes contained long extension products (evident on photographs of the entire gel; data not shown), so the incubations yielded discrete products, but none result from branches in domain 6. These data show that the IVS-E2 products in the D6 mutants are linear and not lariat or broken-lariat RNAs. This conclusion was confirmed by an RNase H experiment that distinguishes between linear and

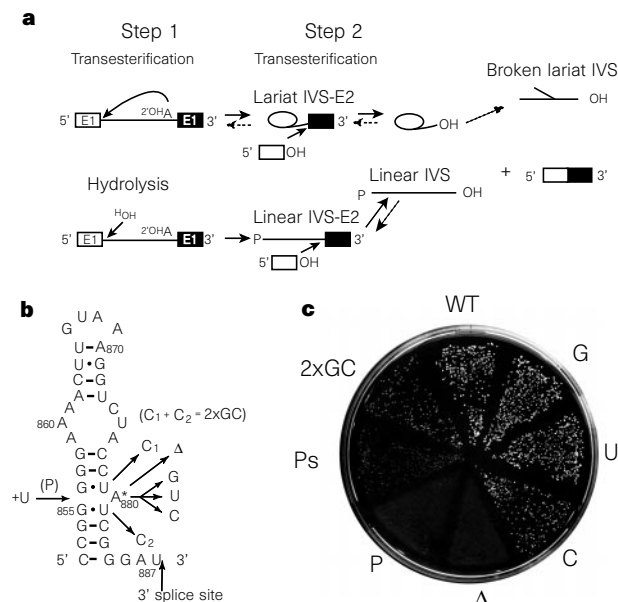


Figure 1 Group II intron splicing pathways and domain-6 mutants. **a**, Splicing pathways of group II introns. The main self-splicing pathways and products for this group II intron^{4,28} are shown; the abbreviations and labels refer to various RNA species in subsequent figures. **b**, Diagram of D6 and mutant alleles. The primary sequence and putative secondary structure of D6 of al5 γ are shown. Nucleotide positions are numbered from the first nucleotide of the intron²⁷. Six mutations affecting the branch nucleotide (A880) of D6 are indicated. **c**, Glycerol growth of mutant and control strains at 18°C. The photograph shows a glycerol-medium Petri dish on which samples of each strain were incubated at 18°C. The photograph shows a glycerol-medium Petri dish on which samples of each strain were incubated at 18°C for 10 days. These strains grow well on this medium at 30 and 36°C (not shown), although the mutants grow detectably slower than the parent strain (based on growth-rate data in liquid glycerol medium; not shown). WT, wild-type.

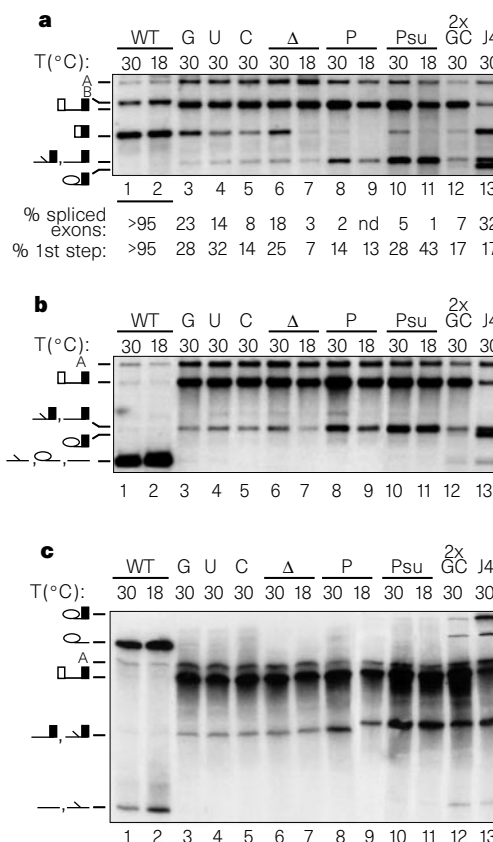


Figure 2 RNA blots showing splicing phenotypes of mutant and control strains. **a**, Exon-containing transcripts of the *COXI* gene. The strains shown were grown in raffinose cultures at the indicated temperatures, and cellular RNA was analysed by RNA blots hybridized with a probe complementary to *COXI* exon 6. Three pre-mRNAs are detected, containing al5 γ alone (diagram), al5 γ and al4 α (band A), and al4 α alone (band B)^{12,21}. Other signals are IVS-E2 and spliced mRNA. The fraction of *COXI* transcripts present as mRNA (percentage of spliced exons) and the fraction of transcripts reacted for the first splicing step (% 1st step = the % spliced exons + %IVS-E2) are reported below each lane. **b**, Intron 5 γ -containing transcripts of the *COXI* gene. Gels of the RNAs analysed in **a** were hybridized with an al5 γ -specific probe. The main bands detected are pre-mRNAs not spliced for al5 γ and band A (see **a**), IVS-E2 and excised IVS. **c**, Branched transcripts. The RNAs above were fractionated on a polyacrylamide gel and hybridized with an al5 γ -specific probe. The locations of intron lariat and linear/broken lariat RNA were defined by products of self-splicing (not shown). T, temperature; WT, wild type.

broken-lariat RNAs (see Supplementary information).

The splicing phenotype of mutant *Psu* shows that first-step hydrolysis can be an efficient reaction *in vivo*. *Psu* is a spontaneous revertant of mutant *P* with an improved ability to grow on glycerol medium at 18 °C as a result of having a dominant nuclear mutation (V.T.C., M.P., P.S.P. and A.M.P., manuscript in preparation). Mutant *Psu* remains inefficient for step 2, but has the highest level of step-1 reaction of this set of D6 mutants (Fig. 2a, lanes 10, 11). It accumulates a high level of IVS-E2 RNA, all of which is shown to be linear by gel mobility (Fig. 2a–c, lanes 10, 11) and reverse transcriptase primer extension (Fig. 3b, c, lane 7). It has previously been noted that branching seems to be important for progression of self-splicing through the second step^{14,15}, and our data extend that inference to living cells. Because the severity of the *in vivo* second-step defect varies widely among these branchpoint mutants, it seems that the absence of branching in the first step is not the main cause of the second-step defects seen here.

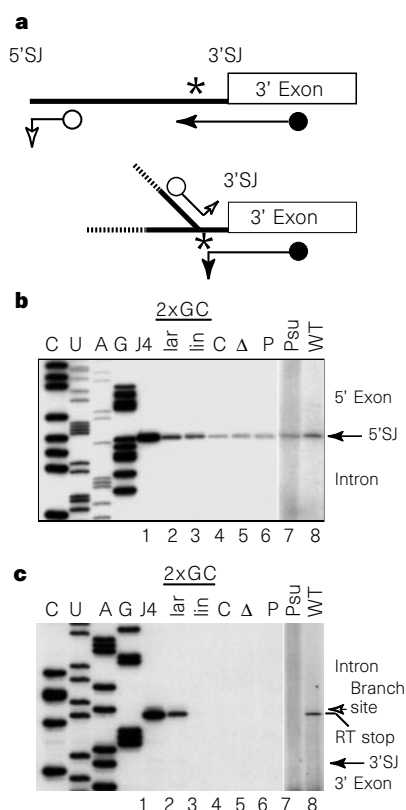


Figure 3 Reverse-transcriptase primer extension analysis of IVS-E2 RNA. **a**, Diagram of reverse transcriptase primer extension experiments. Expected outcomes of reverse transcriptase primer extension experiments using IVS-E2 RNAs and primers complementary to sequences near the 5' or 3' splice junctions (SJ) of the intron (open and filled arrows, respectively). The asterisk indicates the location of the branch nucleotide, at which strong stops will occur with lariat or broken-lariat templates but not with linear templates. **b**, Mapping the 5' end of IVS-E2 RNAs. IVS-E2 RNAs were purified from polyacrylamide gels of RNA from the control strains (wild type and mutant J4), D6 mutants C, Δ , P and 2XGC (see Fig. 1a) and a revertant of the P mutant (*Psu*). Reverse-transcriptase primer extension using the intron-specific primer detected the predicted main product (cDNAs ending at the first nucleotide of the intron). This gel provides an estimate of the amount of IVS-E2 RNA present in these samples. **c**, Mapping the branch site in IVS-E2 RNAs. The IVS-E2 RNAs in **b** were analysed using the downstream exon-specific primer. Here a primer extension product 67 nucleotides long, as in lanes 1, 2 and 8, denotes branching at A880, and the samples in lanes 3–7 did not yield a signal at that position.

The properties of the D6 mutant 2XGC (Fig. 1b), in which the GU wobble pairs flanking the branch nucleotide were changed to GC pairs, show that the hydrolysis and transesterification pathways can coexist *in vivo*. *In vitro*, this mutation has little effect on self-splicing or branching at A880 in high-salt reactions (V.T.C., M.P., P.S.P. and A.M.P., manuscript in preparation); however, this strain splices only about 7% of the pre-mRNA in cells grown at 30 °C (Fig. 2a, lane 12), and reverse-transcriptase primer extension analysis of gel-purified branched and linear IVS-E2 RNAs from the mutant confirm the identification of those RNAs (Fig. 3c, lanes 2, 3).

These results show that mutations blocking group II intron branching allow enough splicing for respiratory growth. We provide strong evidence that linear IVS-E2 RNA, the expected product of first-step splicing by hydrolysis, is present in each branchpoint mutant. The simplest interpretation of our findings is that the linear IVS-E2 RNA is an intermediate that leads to the spliced mRNA found. Other interpretations are not evident without making *ad hoc* assumptions. For example, it is possible that the linear IVS-E2 RNAs present are end products of a side (hydrolysis) pathway activated by these D6 mutations, in which case the spliced mRNAs could result from another pathway, perhaps involving the formation of cryptic or mutant branches. Explanation of the observed instability of excised intron RNA requires the assumption that such intron lariats are efficiently debranched (and then degraded). Disruption of the nuclear *DBR1* gene, which encodes the lariat debranching enzyme¹⁶, did not lead to accumulation of excised intron RNAs, making the above alternative hypothesis unlikely (see Supplementary information).

Our finding that a group II intron can use first-step hydrolysis instead of transesterification in yeast mitochondria has several implications. First, a few group II introns lack an unpaired adenosine in domain 6 (ref. 17), so they may splice *in vivo* mainly by hydrolysis. Second, it is possible that wild-type group II introns can splice *in vivo* by a mixture of branching and hydrolysis reactions, in which case activation or inhibition of the hydrolysis pathway could serve some regulatory purpose. For example, some group II introns engage in transposition events that are initiated by reverse splicing^{18,19}. Cells splicing by the hydrolysis pathway would have low levels of excised intron RNA and so would be relatively deficient for intron mobility. Third, given the possible evolutionary relationship between group II and spliceosomal introns⁴, we wonder whether some spliceosomal introns can carry out the first splicing reaction by hydrolysis instead of transesterification. Finally, it is possible that the ability of this group II intron to carry out first-step splicing by hydrolysis is an ancient feature of a predecessor of group II introns. Perhaps an early version of the first-step reaction centre was limited to the hydrolysis reaction, and the ability to do transesterification evolved later. □

Methods

Mutant alleles, yeast strains, and growth conditions. Site-directed mutations were transformed into mitochondria of a ρ^0 derivative (which completely lacks mtDNA) of strain MCC109 (*MAT α ade2-101 ura3-52 kar1-1*), and the mutant alleles placed in otherwise wild-type mtDNAs by recombination as described^{20,21}. All of these mtDNA mutations were analysed here in the nuclear background of strain ID41-6/161 (*MAT α ade1 lys1*)²². Growth on solid glycerol medium (2% yeast extract, 2% bacto peptone, 4% glycerol, 2% bacto agar) was analysed by streaking portions of glycerol cultures at 30 °C containing about 500 cells on sectors of YPG plates with incubation at 18, 30 and 36 °C. For RNA experiments, cells were grown in liquid raffinose medium (2% yeast extract, 2% bacto peptone, 2% raffinose) to A_{600} of 1–1.5 at 18 or 30 °C, as indicated, with vigorous shaking. This medium was used because it supports relatively high levels of mitochondrial gene expression, but raffinose growth does not depend on respiration.

RNA blots. RNA was prepared from cell homogenates and balanced to $\pm 20\%$ for content of *COXII* mRNA¹². RNAs were displayed on 1.2% agarose²³ or 4% polyacrylamide denaturing gels, blotted and hybridized with ³²P-labelled probes¹¹. We have shown previously that linear and broken-lariat RNAs of

this intron migrate identically on these gels^{2,11,24}. Broken-lariat IVS-E2 RNA was shown to co-migrate with linear IVS-E2 RNA by re-electrophoresis of lariat IVS-E2 RNA from strain J4 alongside RNA from these D6 mutants. Each strain was analysed with at least three independent batches of RNA and representative lanes are shown. Phosphorimager scanning was used to quantify the levels of various RNA signals.

Reverse-transcriptase primer extension. RNAs were fractionated on denaturing polyacrylamide gels (Fig. 2c) and putative lariat or linear IVS-E2 RNAs were recovered from gel slices. Portions were used as templates in reverse-transcriptase primer extension experiments using 5'-end labelled oligonucleotide probes complementary to nucleotides 120–138 of the intron (Fig. 3b) or 39–60 of exon 6 (Fig. 3c)²⁵. Sequence ladders for each panel were made using as template a PCR product containing aI5y and flanking exon sequences amplified from the wild-type yeast strain, and the same oligonucleotide primer used for the reverse-transcriptase primer extensions. The Titan One Tube RT-PCR system was used according to manufacturer's instructions (Boehringer), with oligonucleotide primers in flanking exons to amplify spliced cDNAs as a test of splicing accuracy.

Received 4 September; accepted 10 November 1997.

- Cech, T. R. In *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 239–270 (Cold Spring Harbor Laboratory Press, NY, 1993).
- Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. Group II intron self-splicing: alternative reaction conditions yield novel products. *J. Biol. Chem.* **263**, 3432–3439 (1988).
- Daniels, D., Michels, W. J. & Pyle, A. M. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in-vitro reaction rates and products. *J. Mol. Biol.* **256**, 31–49 (1996).
- Michel, F. & Ferat, J. L. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461 (1995).
- Liu, Q. *et al.* Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. *J. Mol. Biol.* **267**, 163–171 (1997).
- Arnberg, A., Ommen, G. V., Grivell, L., Bruggen, E. V. & Borst, P. Some yeast mitochondrial RNAs are circular. *Cell* **19**, 313–319 (1980).
- Hensgens, L. A. M. *et al.* Variation, transcription, and circular RNAs of the mitochondrial gene for subunit 1 of cytochrome c oxidase. *J. Mol. Biol.* **164**, 35–58 (1983).
- Contrad-Webb, H., Perlman, P. S., Zhu, H. & Butow, R. A. The nuclear *SUV3-1* mutation affects a variety of post-transcriptional processes in yeast mitochondria. *Nucleic Acids Res.* **18**, 1369–1376 (1990).
- Margossian, S., Li, H., Zassenhaus, H. & Butow, R. The DEXH box protein Suv3p is a component of a yeast mitochondrial 3'-to-5' exonuclease that suppresses group I intron toxicity. *Cell* **84**, 199–209 (1996).
- Margossian, S. & Butow, R. RNA turnover and the control of mitochondrial gene expression. *Trends Biochem. Sci.* **21**, 392–396 (1996).
- Peebles, C. L., Belcher, S. M., Zhang, M., Dietrich, R. C. & Perlman, P. S. Mutation of the conserved first nucleotide of a group II intron from yeast mitochondrial DNA reduces the rate but allows accurate splicing. *J. Biol. Chem.* **268**, 11929–11938 (1993).
- Boulanger, S. C. *et al.* Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3' splice site selection. *Mol. Cell. Biol.* **16**, 5896–5904 (1996).
- Costanzo, M. C. & Fox, T. D. Control of mitochondrial gene expression in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **24**, 91–113 (1990).
- Jacquier, A. & Jacquesson-Breuleux, N. Splice site selection and role of the lariat in a group II intron. *J. Mol. Biol.* **219**, 415–428 (1991).
- Chanfreau, G. & Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II introns. *EMBO J.* **12**, 5173–5180 (1993).
- Chapman, K. B. & Boeke, J. D. Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* **65**, 483–492 (1991).
- Michel, F., Umesono, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* **82**, 5–30 (1989).
- Eskes, R., Yang, J., Lambowitz, A. & Perlman, P. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88**, 865–874 (1997).
- Zimmerly, S. *et al.* A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**, 529–538 (1995).
- Butow, R. A., Belcher, S. M., Moran, J. V., Henke, R. M. & Perlman, P. S. Transformation of *S. cerevisiae* mitochondria using the biolistic gun. *Methods Enzymol.* **264**, 265–278 (1995).
- Boulanger, S. C. *et al.* Studies of point mutants define essential nucleotides in the domain 5 substructure of a group II intron. *Mol. Cell Biol.* **15**, 4479–4488 (1995).
- Moran, J. V. *et al.* Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: Initial definition of the maturase domain of the group II intron A12. *Nucleic Acids Res.* **22**, 2057–2064 (1994).
- Goda, S. K. & Minton, N. P. A simple procedure for gel electrophoresis and Northern blotting of RNA. *Nucleic Acids Res.* **23**, 3357–3358 (1995).
- Peebles, C. L. *et al.* A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–223 (1986).
- Podar, M., Dib-Hajj, S. & Perlman, P. S. A UV-induced, Mg²⁺-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA* **1**, 828–840 (1995).
- Perlman, P. & Podar, M. Reactions catalyzed by group II introns in vitro. *Methods Enzymol.* **264**, 66–86 (1995).
- Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A. & Macino, G. Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit I of cytochrome oxidase. *J. Biol. Chem.* **255**, 11927–11941 (1980).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. This work was supported by NIH grants to P.S.P. and A.M.P. M.P. was a fellow of the Robert A. Welch Foundation.

Correspondence and requests for materials should be addressed to P.S.P. (e-mail: perlman@utsw.swmed.edu).

X-ray crystal structure of arrestin from bovine rod outer segments

Joachim Granzin, Ursula Wilden, Hui-Woog Choe*, Jörg Labahn, Bianca Krafft & Georg Büldt

Forschungszentrum Jülich, Institut für Biologische Informationsverarbeitung, D-52425 Jülich, Germany

* Present address: Department of Chemistry, Natural Science College, Chonbuk National University, 560-756 Chonju, S. Korea

Retinal arrestin is the essential protein for the termination of the light response in vertebrate rod outer segments. It plays an important role in quenching the light-induced enzyme cascade by its ability to bind to phosphorylated light-activated rhodopsin (P-Rh*). Arrestins are found in various G-protein-coupled amplification cascades. Here we report on the three-dimensional structure of bovine arrestin (relative molecular mass, 45,300) at 3.3 Å resolution. The crystal structure comprises two domains of anti-parallel β -sheets connected through a hinge region and one short α -helix on the back of the amino-terminal fold. The binding region for phosphorylated light-activated rhodopsin is located at the N-terminal domain, as indicated by the docking of the photoreceptor to the three-dimensional structure of arrestin. This agrees with the interpretation of binding studies on partially digested and mutated arrestin.

Arrestin (also called S-antigen or 48K protein) was first observed in the outer segment of the bovine photoreceptor¹. In dark-adapted rod outer segments the protein freely diffuses in the cytoplasmic space whereas after exposure of the photoreceptor to light it becomes bound to the disk membrane. The light-activated form of the visual pigment rhodopsin (Rh*) interacts with the retinal G protein transducin ($T_{\alpha\beta\gamma}$), thereby initiating the exchange of a GDP molecule for GTP at the α -subunit of transducin² (T_{α}). In its GTP-binding form transducin dissociates from Rh*, and two T_{α} subunits activate one effector molecule, a cyclic GMP phosphodiesterase (PDE)^{3,4}, by binding to its two inhibitory subunits PDE γ . The result is a rapid decline in the concentration of the internal transmitter cyclic GMP. Because the interaction of Rh* and transducin takes only about 1 ms (ref. 5), a single Rh* can subsequently interact with hundreds of transducin molecules. The turnover number for PDE can be in the order of several thousand hydrolysed cGMP per PDE per second. Thus for a limitation of the light response as well as for a

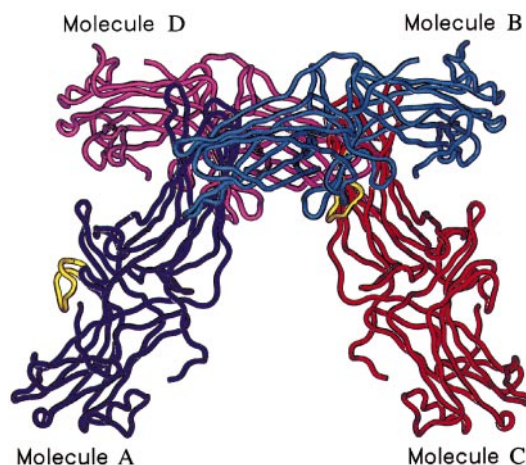


Figure 1 C α -backbone of the four molecules (A, B, C, D) showing their packing in the asymmetric unit. The loop between residues 68 to 78 in molecules A and B is shown as a yellow ribbon.

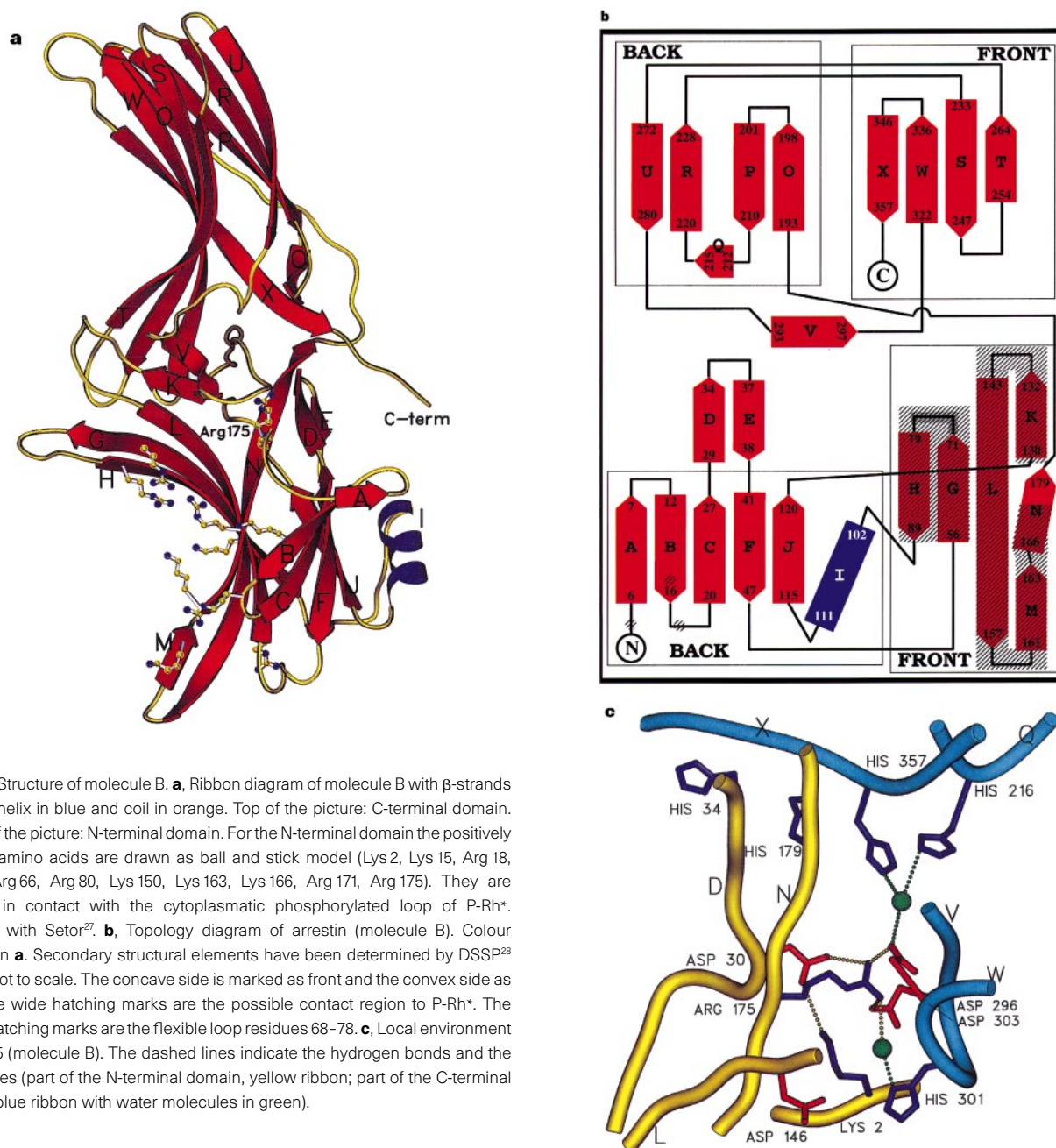
fast recovery, a rapid and effective elimination of Rh* is essential, before it can activate too many PDE molecules. This inactivation of Rh* is accomplished in two steps: phosphorylation of Rh* reduces its ability to catalyse the nucleotide exchange of transducin⁶⁻⁸, and subsequent binding of arrestin to P-Rh* completely shields it from further interaction with transducin⁸⁻¹⁰.

Arrestin was prepared and purified using its light-dependent binding to P-Rh* from bovine retinae¹¹. Phases for a room-temperature (RT) structure were determined by MIRAS (multiple isomorphous replacement with anomalous scattering) (see Table 1). At cryotemperature ($T = 100\text{K}$) better quality data were obtained only for native crystals. The low-temperature data were not isomorphous with the room-temperature data. Therefore, it was necessary to project the density from the RT data in the molecule mask onto the unit cell of the low-temperature data using the program MAVER¹². The low-temperature structure was subsequently refined by XPLOR¹³.

The monomer has dimensions of roughly $90 \times 39 \times 50 \text{ \AA}^3$. The asymmetric unit contains four molecules (Fig. 1), each having a

length of 404 amino acids¹⁴, which are related by a set of non-crystallographic symmetry (NCS) operators. The interpretability of the carboxy-terminal electron density differs remarkably. The densities of molecules A and C are truncated at residue 368 and for molecules B and D at residue 363 (see Methods). Each of the two domains fold in two layers of antiparallel β -sheet on top of each other (Fig. 2a). The N-terminal domain sandwich is formed by the β -sheets A:B:C:F:J and N:L:G:H, the C-terminal domain by β -sheets U:R:P:O and T:S:W:X (Fig. 2b). The interdomain linkage can be described as a hinge region. Additionally the N-terminal domain contains a short α -helix (in Fig. 2a, b, labelled with I) on the convex side.

Note that short intermolecular β -barrels are formed by molecule A (β -strand O) with molecule D (β -strand H) and molecule C (β -strand O) with molecule B (β -strand H). The molecules are equivalent in pairs, A, C and B, D, respectively. The root-mean-square (r.m.s.) deviation for all atoms between A and C is equal to $0.11 \text{ \AA}$ and between B and D to $0.023 \text{ \AA}$. Molecules belonging to non-equivalent pairs differ from each other significantly in the



loop spanning residues 68–78 (maximum r.m.s. difference, 23 Å; see Fig. 1), at the interface between hinge and residues 181–201 (maximum r.m.s. difference, 12.5 Å) of the C-terminal domain, at residues 262–272 (maximum r.m.s. difference, 4 Å) and at the highly flexible C-terminal end starting with residue 361 (maximum r.m.s. difference 6.5 Å). An additional remarkable deviation is indicated by the mean temperature factor (B -factor averaged over all protein atoms; see Methods). The B -factor for molecules A and C and for the C-terminal domain of molecules B and D is in the range of about 20 Å², respectively, but the N-terminal domain of molecules B and D has an average value of about 46 Å². The two non-equivalent pairs of molecules result from different packing environments. Their conformations may result in different interactions with P-Rh*.

From binding studies on partially digested and mutated arrestin^{15–17} it was inferred that the phosphorylation-recognition domain should be located predominantly between residues 158 and 185. This sequence includes motif III, one of five highly conserved motifs defined for all known arrestin sequences by a protein fingerprint database, PRINTS¹⁸ (Fig. 3). In particular, Arg 175 (β-strand N) has been predicted to play two key roles: to control the transition into a high-affinity binding state and to serve as a phosphorylation-sensitive trigger^{16,17}.

The N-terminal segments (Fig. 2a, b) spanning residues 56–91 (includes motif II), 129–175 (part of motif III) as well as the single residues Lys 2, Lys 15, Arg 18, can be involved in a possible contact to the cytoplasmic side of P-Rh*. The residues are mostly positioned at the concave dome-side. This roof has a rough diameter of 35 Å and a depth of 24 Å. The accessible domain is covered by 11 positively charged amino acids (see legend to Fig. 2a) and six negatively charged amino acids (Asp 82, Glu 148, Asp 158, Glu 160, Glu 161, Asp 162). Two pairs of side chains form salt bridges (Arg 80–Asp 82; Arg 171–Glu 148). At Lys 176 the twisted β-strand N leaves the concave domain surface which prevents any of the following amino

acids from interacting with the C-terminal phosphorylated sites of Rh*. In its actual conformation Arg 175 forms a double salt bridge to Asp 30 and Asp 296 and hydrogen-bonds to Asp 303 and one water molecule (Fig. 2c). If Arg 175 actually serves as a phosphorylation-sensitive trigger and controls the transition into a high-affinity binding state^{16,17}, then the salt bridges and the hydrogen bonds must be broken. This would result in an exposed Arg 175 in the highest point of the dome. The residues Lys 176 to Gln 178 (end of β-strand N) are positioned in the inner hinge region and are not accessible. The backbone of Pro 181 to Gly 185 is located on the convex dome-side (see Fig. 2a, b) and belongs to the C-terminal domain. This structural picture suggests that the phosphorylation-recognition region extends only to Arg 175 and not to residue 185, as proposed the first time by Palczewski *et al.*¹⁵. Although residues 176 to 185 have no such role, their location in the hinge region indicates a possible involvement in a trigger function. Participation of histidine residues (Fig. 2c) has not been noted up to now.

Molecules A and C are in a so-called 'closed conformation', because the loop between 68 and 78 (motif II; between β-strands G and H; see Fig. 1 molecule A, the yellow ribbon) prevents in molecules A and C the C-terminal cytoplasmic loop of P-Rh* from binding to the supposed site of arrestin. In molecules B and D this flexible loop adopts an 'open conformation' with undisturbed

Motif I ²⁸KRDYIDHVERVEPVDGVVLVDPE⁵⁰
 Motif II ⁶⁵FRYQEDIDVMGLSFRRDL⁸³
 Motif III ¹⁶¹EDKIPKSSVRLIRKVQ¹⁷⁸
 Motif IV ²⁸⁶NNRERRGIALDGKIKHEDT³⁰⁴
 Motif V ³⁷²DENFVFEEFARQNLK³⁸⁶

Figure 3 The five fingerprint regions based on the highly conserved motifs defined for all known arrestin sequences.

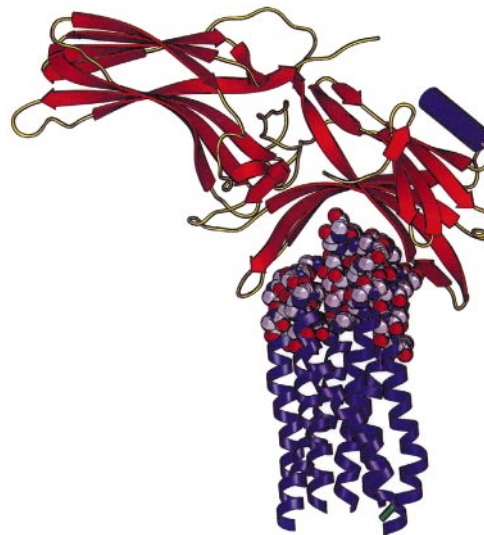


Figure 4 Possible coordination of arrestin (molecule B) and the receptor (theoretical model of metarhodopsin II²³, Protein Data Bank id: 1boj, entry without loops). The cytoplasmic loops are reconstructed with the loop database of the program O²⁶ and shown in cpk-mode. The dockings were made manually without energy minimization. The membrane-spanning helices are drawn as ribbons.

Table 1 Crystallographic data

Data collection and phasing statistic			Derivatives (RT)		
Data set	Native 1 (RT)	Native 2 ($T = 100\text{K}$)	HgCl ₂ (1 mM per 6 h)	K ₂ PtCl ₄ (0.5 mM per 20 h)	K ₂ ReCl ₆ (1 mM per 2 h)
Resolution (Å)	20–3.45	130–3.3	20–3.79	20–5.0	24.8–6.8
Completeness (%)	82.4(61.1)*	95.1(79.7)*	82.6	79.5	73.7
Redundancy	3.6	4.9	3.6	4.6	4.9
R_{merge}	0.094(0.180)*	0.075(0.172)*	0.072	0.080	0.077
R_{iso}			0.308	0.229	0.157
Binding sites			8	6	5
Phasing power			1.89	1.03	0.94
R_{cullis}			0.56	0.83	0.94
Anomalous data			yes	no	no
Mean figure of merit	0.416				

$R_{\text{merge}} = \sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)$, where $I_i(h)$ is the i th measurement and $\langle I(h) \rangle$ is the mean of all measurements of $I(h)$.

$R_{\text{iso}} = \sum_i |F_P - F_{PH}| / \sum_i F_P$, where F_P and F_{PH} are the structure factor amplitudes of the native and derivative data, respectively.

Phasing power (isomorphous) = $\langle |F_{PH}| \rangle / \langle |F_P| \rangle$, where $\langle |F_{PH}| \rangle$ is the mean calculated amplitude for the heavy-atom model and ϵ is the lack of closure error.

$R_{\text{cullis}} = \sum_i |F_{PH} \pm F_P| - F_{\text{Hcal}} / \sum_i |F_{PH} \pm F_P|$ defined for centric reflections, where $|F_P|$, F_{PH} and F_{Hcal} are native, derivative and calculated heavy-atom structure factor amplitudes, respectively.

(*) Indicates the value for the outermost shell.

electron density. In this conformation the molecules seem to be in a proper binding state because the loop spanning residues 68–78 display no steric hindrance for an interaction with the C-terminal phosphorylated loop of Rh*. The closed conformation may indicate an inactive state. It prevents an interaction between the postulated phosphorylation-recognition domain of arrestin and P-Rh* (ref. 19), which would assign a switch function to this loop. On the other hand the flexibility of this loop may be used to tighten P-Rh* binding.

Comparing the structure of arrestin with the contents of the FSSP database²⁰ eight proteins with a Z-score (strength of structural similarity in standard deviation) greater than 6.0 were found. They are all dominated by β -strands (coagulation factor XIII, Z-score 7.7; sialidase, Z-score 7.2; PapD, Z-score 6.9; human vascular cell adhesion molecule-1, Z-score 6.8; chitinase A, Z-score 6.7; fragment of human fibronectin, Z-score 6.5; tenascin, Z-score 6.3; MHC class II i-ek, Z-score 6.3) but there are no obvious functional homologies. Structural similarities between arrestin and transducin were not observed^{14,21,22}.

Some other reported possible functional segments of arrestin^{16,17} are just speculative without the three-dimensional data of the C-terminal phosphorylated loop of Rh*, but the docking picture (Fig. 4) suggests that the N-terminal domain of arrestin can shield the cytoplasmic loops of rhodopsin completely from transducin. The conformation of the C-terminal roof is more than 10 Å smaller and the positively charged environment is completely surrounded by glutamates, which should prevent rhodopsin from binding to this domain. It is not possible to conclude from the arrestin–P-Rh* model²³ in which way arrestin discriminates between the phosphorylated C-terminal tail and the other potentially phosphorylated sites on the cytoplasmic side of P-Rh*.

On the basis of the X-ray structure in combination with data from the literature^{15–17} it is possible to locate the predicted phosphorylation-recognition domain (see Fig. 2b, wide hatching), which interacts with the phosphorylated C-terminal loop region of P-Rh*. From the three-dimensional arrangement of the relevant amino acids, the following steps are conceivable: in the case of a closed conformation the recognition segment (motif III) has to become accessible by the conformational change of loop 68 to 78 (motif II). Upon binding of P-Rh* the salt bridges in the hinge region (motif I and IV), especially Arg 175, probably will be broken. Subsequent conformational changes of motif II may tighten the binding of P-Rh*. □

Methods

Data collection. Arrestin was purified and crystallized as described²⁴. Native and heavy-atom derivative data were recorded from room-temperature crystals with a MAR Research image-plate area detector at different Synchrotron radiation sources (Instrument X11 at EMBL outstation at the Deutsches Elektronen Synchrotron, Hamburg, Germany; PX7.2/PX9.5 at SERC Daresbury Laboratory, GB; DW21B at LURE Orsay, France). Crystal dimensions were around $0.3 \times 0.2 \times 0.15$ mm³. One to two crystals were used per data set (Table 1). The final structure refinement was done on cryo-cooled crystals (cryo-protectant as reported²⁴). The space group is $P2_12_12$ with $a = 171.22$ Å, $b = 188.77$ Å, $c = 90.79$ Å at room temperature and $a = 169.17$ Å, $b = 185.53$ Å, $c = 90.93$ Å at cryo-temperature ($T = 100$ K) with 4 molecules per asymmetric unit.

Structure solution. The structure was solved by multiple isomorphous replacement with anomalous scattering (Table 1). All heavy-atom sites were located in a difference Patterson map and origin correlated by cross-difference Fouriers. The non-crystallographic symmetry (NCS) operators were determined from heavy-atom derivatives. Heavy-atom positions were refined and phases to 3.79 Å (at room-temperature RT) were calculated using MLPHARE²⁵. The initial MIRAS-density was improved and extended to 3.45 Å (RT) by solvent flattening (solvent content of 68.5% by volume), histogram matching and NCS averaging using DM²⁵. Map interpretation and model building were done using the program O²⁶. To improve the resolution we transformed the electron density map by MAVER¹² to the cryo-cooling data (see above). Using

readjusted NCS operators the phases were extended to 3.3 Å by DM. The structure was initially refined with NCS constraints. The NCS constraints were relaxed during the refinement progress (torsion angle dynamics, X-PLOR¹³). The temperature factors were restrained in groups of main- and side-chain atoms. Solvent mask correction and overall anisotropic temperature factor applied to F_0 improved the refinement. The final model contains residues 1–368 of molecules A and C and residues 1–363 for molecules B and D. The four molecules in the asymmetric unit have 285 water molecules added. The R -factor of the current structure is $R = 25.4\%$ with a free R -factor of 31.6% (the R -factor calculations were done with no sigma cutoff; the test set contains 5% of the data). The r.m.s. deviations from ideality are 0.006 Å for bond lengths and 1.233° for bond angles. Of the non-glycine residues, none is within the energetically disallowed region of the Ramachandran plot.

Biochemical checks. The deficiency of the amino acids at the C-terminal end was examined by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) and western blot (different antibodies, data not shown). The crystallized protein has the full amino acid length.

Received 23 September; accepted 20 November 1997.

- Kühn, H. Light-regulated binding of rhodopsin kinase and other proteins to cattle photoreceptor membranes. *Biochemistry* **17**, 4289–4395 (1978).
- Fung, B. K.-K. & Stryer, L. Photolyzed rhodopsin catalyzes the exchange of GTP for bound GDP in retinal rod outer segments. *Proc. Natl Acad. Sci. USA* **77**, 2500–2504 (1980).
- Fung, B. K.-K., Hurley, J. B. & Stryer, L. Flow of information in the light-triggered cyclic nucleotide cascade of vision. *Proc. Natl Acad. Sci. USA* **78**, 152–156 (1981).
- Uchida, S., Wheeler, G. L., Yamazaki, A. & Bitensky, M. W. A GTP-protein activator of phosphodiesterase which forms in response to bleached rhodopsin. *J. Cyclic Nucl. Res.* **7**, 95–104 (1981).
- Vuong, T. M., Chabre, M. & Stryer, L. Millisecond activation of transducin in the cyclic nucleotide cascade of vision. *Nature* **311**, 659–661 (1984).
- Wilden, U. & Kühn, H. Light-dependent phosphorylation of rhodopsin: number of phosphorylation sites. *Biochemistry* **21**, 3014–3022 (1982).
- Miller, J. L., Fox, D. A. & Litman, B. J. Amplification of phosphodiesterase activation is greatly reduced by rhodopsin phosphorylation. *Biochemistry* **25**, 4983–4988 (1986).
- Wilden, U., Wüst, S. W. & Kühn, H. Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kDa protein of rod outer segments. *Proc. Natl Acad. Sci. USA* **83**, 1174–1178 (1986).
- Wilden, U. Duration and amplitude of the light-induced cGMP hydrolysis in vertebrate photoreceptors are regulated by multiple phosphorylation of rhodopsin and by arrestin binding. *Biochemistry* **34**, 1446–1454 (1995).
- Schleicher, A., Kühn, H. & Hofmann, K. P. Kinetics, binding constant, and activation energy of the 48-kDa protein-rhodopsin complex by extra-metorhodopsin II. *Biochemistry* **28**, 1770–1775 (1989).
- Wilden, U., Wüst, S. W., Weyand, I. & Kühn, H. Rapid affinity purification of retinal arrestin (48kDa protein) via its light-dependent binding to phosphorylated rhodopsin. *FEBS Lett.* **207**, 292–295 (1986).
- Kleywegt, G. J. Making the most of your search model. *CCP4/ESF-EACMB News. Prot. Crystallogr.* **32**, 32–36 (1996).
- Brunger, A. T. *X-PLOR, Version 3.1-3.851, A System for X-ray Crystallography and NMR* (Yale University, New Haven, CT, 1992–1997).
- Shinohara, T. et al. Primary and secondary structure of bovine retinal S antigen (48-kDa). *Proc. Natl Acad. Sci. USA* **84**, 6974–6979 (1987).
- Palczewski, K., Pulvermüller, A., Buczylo, J. & Hofmann, K. P. Phosphorylated rhodopsin and heparin induce similar conformational changes in arrestin. *J. Biol. Chem.* **266**, 18649–18654 (1991).
- Gurevich, V. V. & Benovic, J. L. Visual arrestin interaction with rhodopsin. *J. Biol. Chem.* **268**, 11628–11638 (1993).
- Gurevich, V. V. & Benovic, J. L. Mechanism of phosphorylation-recognition by visual arrestin and the transition of arrestin into high affinity binding state. *Mol. Pharmacol.* **51**, 161–169 (1997).
- Attwood, T. K. et al. Novel developments with the PRINTS protein fingerprint database. *Nucleic Acids Res.* **25**, 212–216 (1997).
- Xinyu, Z., Palczewski, K. & Ohguro, H. Mechanism of rhodopsin phosphorylation. *Biophys. Chem.* **56**, 183–188 (1995).
- Holm, L. & Sander, C. The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.* **22**, 3600–3609 (1994).
- Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin- α complexed with GTP γ S. *Nature* **366**, 654–666 (1993).
- Lohse, M. J., Benovic, J. L., Codina, J., Caron, M. G. & Lefkowitz, R. J. β -Arrestin: a protein that regulates β -adrenergic receptor function. *Science* **248**, 1547–1550 (1990).
- Pogozheva, I. D., Lomize, A. L. & Mosberg, H. I. The transmembrane 7- α -bundle of rhodopsin: distance geometry calculation with hydrogen bonding constraints. *Biophys. J.* **70**, 1963–1985 (1997).
- Wilden, U., Choe, H.-W., Krafft, B. & Granzin, J. Crystallization and preliminary X-ray analysis of arrestin from bovine rod outer segment. *FEBS Lett.* **415**, 268–270 (1997).
- Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Evans, S. V. SETOR: hardware lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).
- Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

Acknowledgements. We thank A. Cousin and R. Esser for the arrestin protein preparation from bovine eyes; L. A. Donoso for the arrestin antibodies; U. B. Kaupp for continuous support of the project; K.-P. Hofmann and M. J. Lohse for helpful discussions. This work was funded by a grant from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to J.G. (e-mail: J.Granzin@fz-juelich.de). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession number 1ayr).

At the flick of a molecular switch

This week's product update features tools for the cell biologist, including electroporation and transfection systems, transcription and translation kits, new fluorescent markers and reagents for neuroscience.

Modulis

From MBI Fermentas

An in vitro transcription and translation system for RNA and protein generation

This system consists of single-tube reactions that can be used separately to generate RNA from a DNA template and protein from an RNA template. Together, they can create a linked system that yields protein in as little as 45 minutes. In this protocol, the DNA to be transcribed is added to the transcription mix followed by either SP6 or T7 transcriptzyme mix, and a 15-min incubation step. Then, with no purification required, an aliquot of the transcription products is added to the rabbit reticulocyte translation mix (provided), and this is incubated for 30 minutes. Products of the transcription reaction may be frozen for use in subsequent translation reactions. Moreover, RNA from other sources can be translated by using the adaptor buffer provided.

Reader Enquiry No. 100

FuGene 6

From Boehringer Mannheim

A high-efficiency, low-toxicity transfection reagent for use with or without serum

This system uses a non-liposomal formulation with minimal cytotoxicity to produce high levels of transfection across many eukaryotic cell lines (with or without serum). The procedure is simple — the user mixes the reagent with DNA and applies it to the cells, allowing 10–15 min of incubation. Owing to the low toxicity, it is not necessary in many cases to change the medium until a reporter gene assay is performed.

Reader Enquiry No. 101

Dynabeads

From Dynal

Magnetic anti-epithelial cell beads for separation of rare circulating epithelial tumour cells from peripheral blood and bone marrow

These beads are stated to enrich viable target tumour cells from mononuclear cells by more than a thousandfold in less than 1 h, allowing the investigator to screen larger sample volumes. Cells are separated based on the recognition of a glycopeptide by monoclonal Ber-EP4-coated beads. A low-cost magnet is used to collect the beads for immunocytochemistry and is stated to yield four times as many positive cells per slide. This enrichment method can also be used for PCR and RT-PCR procedures.

Reader Enquiry No. 102

Electroporation

EC 100 electroporator

From Sigma

A push-button unit for the transformation of bacterial, yeast, mammalian and plant cells

Designed to facilitate the rapid processing of samples, this electroporator features three pre-optimized transfection settings, a pulse-recharge time of less than 10 s and simple push-button operation. There are two settings for bacteria and yeast cells, using 1-mm and 2-mm cuvettes, and a third for mammalian cells, using 4-mm cuvettes. Features include an audible tone and indicator lamp, which are activated when the pulse has been delivered, a small size that takes up less than 464 cm² of bench space and weighs only 5 kg, the ability to operate from either 120 V AC or 240 V AC power sources and, for convenience, the flexibility to accommodate most standard cuvettes.

Reader Enquiry No. 103

FlexWave

From Bio-Rad

A new flexible wave form technology for electroporating cells

With this technology, cell membranes are permeabilized with electrical pulses that are a combination of radio frequency and square waveforms; whereas, standard electroporation techniques use traditional square-wave and capacitor-discharge systems to provide the electrical pulse. This allows the user to generate flexible wave shapes and pulse patterns not available with traditional systems. This is said to result in improved cell viability, as well as using less DNA sample for the transformation process. The system has a number of ways in which the pulse output can be tailored to suit different cell types, which means that each transformation can be optimized individually.

Reader Enquiry No. 104

Fluorescence

Filter cubes for GFP

From Olympus

Four filter pairs for green fluorescent protein (GFP)-specific applications, using universal infinity-corrected objectives (UIS)

Depending on the form of GFP being used, the autofluorescence behaviour of the specimen and the spectral output of the illumination can be optimized. GFP is rapidly becoming a valuable fluorescent marker for real-time *in vivo* monitoring of



Drosophila blastoderm captured with an IX70 microscope and the U-M41001 bandpass filter 4.

gene expression and protein localization in cells. It can also be used for lineage analysis and monitoring growth and propagation in embryology. Moreover, GFP does not damage host cells and acts as an *in vivo* marker, growing with the cell, as well as being passed on to daughter cells. In all, Olympus provides more than 30 standard cubes for applications from ratio imaging to fluorescence *in situ* hybridization.

Reader Enquiry No. 105

Anti-green fluorescent protein

From MBL

An antibody purified from mouse ascites fluid using protein-A Sepharose

This anti-GFP monoclonal antibody can detect GFP and its variants in immunoblotting and immunoprecipitation, as well as allowing the detection of GFP-tagged proteins in western blotting. The antibody comes lyophilized and is simply reconstituted in 100 µl of distilled water (IgG concentration: 1.0 mg ml⁻¹ in PBS (pH 7.2), 1% sucrose and 0.1% sodium azide). The lyophilized antibody is stable for five years from the date of manufacture.

Reader Enquiry No. 106

Yellow and blue GFP variants

From Clontech Laboratories

Enhanced yellow and blue fluorescent protein (EYFP and EBFP, respectively), variants of the Aequorea victoria GFP

EYFP contains four amino acid substitutions that shift emission from green (509 nm) towards yellow (527 nm). EYFP- and EGFP-expressing cells can be easily separated by flow cytometry using standard argon laser excitation at 488 nm. The brightness of EYFP is roughly equivalent to that of EGFP, making it well suited to dual-label flow cytometry studies. In addition, the EYFP gene has been codon optimized for high expression in mammalian and

plant cells. The EBFP variant, which is stated to be 2–3 times brighter than other blue variants, also contains four amino acid substitutions that shift emission from green (509 nm) to blue (440 nm), and is said to enhance the brightness, minimize photobleaching and improve solubility of the protein. The EBFP coding sequence has been optimized for human codon usage, providing improved expression in mammalian cells.

Reader Enquiry No. 107

F-2000

From Hitachi

A rapid-scan fluorescence spectrophotometer for self-contained intracellular cation measurement

The use of a plug-in ROM board enables the F-2000 to be used for intracellular cation measurements, without the need for a personal computer. This instrument should prove to be useful for the non-invasive study of cell functions, allowing simultaneously fluorescence monitoring of free or bound *in situ*, cation-specific fluorescent probes. All parameters can be set from the control panel of the instrument and results displayed on the integral monitor. Emission measurements can be made for up to two different fluorescent probes at two different excitation wavelengths for each probe. For example, by using the appropriate fluorescent indicators, simultaneous measurement of cation concentration variations and pH changes can be made. The on-board software allows automatic acquisition of time- and wavelength-based excitation/emission spectra and three methods of intracellular cation concentration calculation, including ratio calculations. Samples can be measured at controlled temperatures using a thermostatic cell holder, which has a stirring and a microsampling assembly that can inject reagents from a microsyringe into the sample cell without opening the sample compartment cover.

Reader Enquiry No. 108

Fluoroskan Ascent FL

From Labsystems

A two-in-one fluorometer and luminometer that improves performance and ease of use

Combined with Ascent software, this instrument covers the full range of fluorometric, glow and flash luminometric applications. It can be used for Ca^{2+} measurement, cell proliferation, cytotoxicity and reporter gene and hybridization assays. Fluoroskan Ascent FL incorporates direct illumination optics that have a highly focused light beam for fluorometric measurements, helping to prevent cross-talk and ensure accurate readings. The choice of two beam-diameter settings (3 mm and 1.5 mm) allows 1- to 96-well

and 384-well readings to be taken. The system has improved its sensitivity through the use of fibreless optics and can also be equipped with filters for luminometric applications. Three reagent dispensers can be fitted onto the instrument, allowing simultaneously the dispensing and measurement of flash luminescence reactions, fluorometric Ca^{2+} measurements and other rapid kinetic applications. For assays requiring precise temperature control, there is an on-board incubator, as well as built-in orbital shaking.

Reader Enquiry No. 109

Neuroscience reagents

Intracellular Ca^{2+} activator

From RBI

This activator is offered with the permission of the Mayo Foundations for Medical Education and Research

This potent activator (NAADP, adenosine 5'-(trihydrogen diphosphate), 2'-(dihydrogen phosphate), P'(5')-3-carboxy-1- β -D-ribofuranosylpyridinium inner salt) signals the release of intracellular Ca^{2+} through an IP_3 - and cyclic ADP ribose-independent mechanism. It is available as a 250-g unit. Over 1,600 products are covered in the company's catalogue, including those for neuroscience, signal transduction and the peripheral nervous system. Other research reagents include adenosines/purinergics, cholinergics, cholecystokinin, dopaminergics, enzyme inhibitors, excitatory amino acids, imidazole binding sites, ion channel modulators, neuropeptides/neuropeptide Y, opioids, neurotoxins and serotonergics.

Reader Enquiry No. 110

Glutamate receptor antibody

From Chemicon

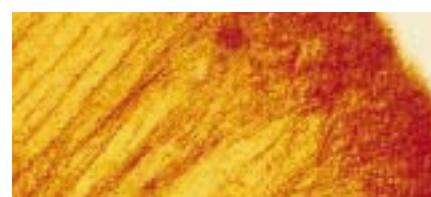
A glutamate receptor 2-specific (GluR2) polyclonal antibody

At least four classes of receptors mediate glutamate's action in the CNS. The major class of receptor, the AMPA receptor, is comprised of four distinct subunits. This new rabbit polyclonal antibody, which is specific to GluR2, is affinity purified and is said to show no reactivity to GluR1, GluR3 or GluR4. It may be used for immunohistochemistry, western blot and immunoprecipitation.

Reader Enquiry No. 111



Chemicon's antibody targets GluR2 receptors.



Immunolabelling of 5-HT2A receptors in rat cortex with Incstar's antibody.

Serotonin-related antibodies

From Incstar

A number of histochemical antisera related to serotonin research

Serotonergic neurons are important mediators of a variety of biological activities, including those in the brain and spinal cord, the peripheral nervous system, the gastrointestinal tract, as well as other sites. Serotonergic influences have been documented in virtually every activity of the central nervous system. Incstar's newest additions to the serotonin product line include antibodies against the 5-HT5A receptor, 5-HT7 receptor and 5-hydroxytryptophan (5-HTP). These new products are a complement to the company's existing antibodies against serotonin, 5-hydroxyindoleacetic acid (5-HIAA), serotonin transported, and the 5-HT2A receptor. All Incstar histochemical antisera are for research use only and not for use in diagnostic procedures.

Reader Enquiry No. 112

Assorted assays and agents

Biotrak cAMP Direct

From Amersham Pharmacia Biotech

A new SPA-based screening assay that permits the direct measurement of intracellular and/or total cyclic AMP

Through the use of cell lysis reagents, the new system eliminates the lengthy extraction steps found in many assays. The protocol is completed by simply adding lysis reagents to

ADVERTISEMENT

Peptides
\$18/residue

Antibodies
monoclonal - \$8500
polyclonal - \$1195

Complete services for your custom peptide and polyclonal or monoclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 205-536-9016

READER ENQUIRY NO. 87